

Univerzita Karlova v Praze  
Matematicko-fyzikální fakulta

## **BAKALÁŘSKÁ PRÁCE**



Matúš Baniar

### **Modelování inflace**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Jitka Zichová Dr.

Studijní program: Matematika

Studijní obor: Finanční Matematika

Praha 2011

Rád by som na tomto mieste poďakoval vedúcej mojej bakalárskej práce RNDr. Jitke Zichovej Dr. za cenné pripomienky a návrhy, za poskytnutie literatúry a softwaru a predovšetkým za jej čas strávený pri konzultáciách.

Prehlasujem, že som svoju bakalársku prácu vypracoval samostatne a výhradne s použitím citovaných prameňov, literatúry a ďalších odborných zdrojov.

Beriem na vedomie, že sa na moju prácu vzťahujú práva a povinnosti vyplývajúce zo zákona č. 121/2000 Sb., autorského zákona v platnom znení, najmä skutočnosť, že Univerzita Karlova v Prahe má právo na uzavretie licenčnej zmluvy o použití tejto práce ako školského diela podľa § 60 odst. 1 autorského zákona.

V Prahe dňa 1.8.2011

Matúš Baniar

**Názov práce:** Modelování inflace

**Autor:** Matúš Baniar

**Katedra:** Katedra pravděpodobnosti a matematické statistiky

**Vedúci bakalárskej práce:** RNDr. Jitka Zichová Dr., Katedra pravděpodobnosti a matematické statistiky

**Abstrakt:** Inflácia, rast všeobecnej cenovej hladiny, je bežným ekonomickým javom, ktorý predstavuje makroekonomický problém. Práca sa zaoberá metódami, pomocou ktorých je možné infláciu modelovať, a teda pochopiť jej vývoj. V prvom rade ide o korelačnú a regresnú analýzu, ktoré sa zaoberajú vzťahom dvoch a viacerých premenných a následným zvolením vhodného matematického modelu. Spracovaný je model lineárnej regresie a metódy, pomocou ktorých analyzujeme jeho adekvátnosť. Ďalšou spracovanou metódou je analýza jednorozmerných časových rád, ku ktorej pristupujeme takzvanou Boxovou-Jenkinsovou metodológiou. Oba prístupy sú ilustrované na skutočných finančných dátach pomocou softwaru Wolfram Mathematica 8.

**Kľúčové slová:** inflácia, korelačná analýza, regresná analýza, časová rada

**Title:** Inflation modeling

**Author:** Matúš Baniar

**Department:** Department of probability and mathematical statistics

**Supervisor:** RNDr. Jitka Zichová Dr., Department of probability and mathematical statistics

**Abstract:** Inflation, the growth of the general price level, is a common economic phenomenon, which is a macroeconomic problem. The thesis deals with the methods by which it is possible to model inflation and therefore to understand its development. In the first case, the correlation and regression analysis, which deal with the relationship of two or more variables and the following selection of the appropriate mathematical model. The model of linear regression is described also with methods by which we analyze its adequacy. Another described method is the analysis of one-dimensional time series, which we apply so called Box–Jenkins methodology. Both approaches are illustrated on real financial data using the software Wolfram Mathematica 8.

**Keywords:** inflation, correlation analysis, regression analysis, time series

# Obsah

<b>Úvod</b> .....	<b>1</b>
<b>1 Základné pojmy</b> .....	<b>2</b>
1.1 Indexové čísla .....	2
1.2 Inflácia .....	3
1.3 RPI v Českej republike .....	3
1.4 Zaznamenávanie inflácie Českým štatistickým úradom .....	4
<b>2 Korelačná analýza</b> .....	<b>6</b>
2.1 Korelačný koeficient.....	6
2.2 Koeficient mnohonásobnej korelácie .....	7
2.3 Koeficient parciálnej korelácie .....	8
<b>3 Regresná analýza</b> .....	<b>10</b>
3.1 Lineárna regresia.....	10
3.2 Koeficient determinácie.....	11
3.3 Normálny model lineárnej regresie .....	11
3.4 Testy nulovosti parametrov .....	12
3.5 Autokorelovanosť rezíduí.....	12
3.6 Multikolinearita.....	14
<b>4 Analýza časových rád</b> .....	<b>16</b>
4.1 Autokorelačná a parciálna autokorelačná funkcia.....	16
4.2 Základné modely MA, AR a ARMA .....	17
4.3 Identifikácia modelu .....	19
4.4 Odhad parametrov modelu .....	20
4.5 Overovanie modelu.....	20
<b>5 Aplikácia</b> .....	<b>22</b>
5.1 Príklad 1 .....	22
5.2 Príklad 2 .....	32
<b>Záver</b> .....	<b>37</b>
<b>Literatúra</b> .....	<b>38</b>
<b>Prílohy</b> .....	<b>39</b>

# Úvod

Táto práca popisuje niektoré základné metódy, ktoré je možné použiť na modelovanie rôznych ekonomických veličín, medzi ktoré patrí aj inflácia.

V prvej kapitole sú definované základne pojmy súvisiace s indexovými číslami a infláciou. Spomenuté sú niektoré základné indexy a metódy na výpočet miery inflácie pre dané obdobie. V závere kapitoly je popísané zaznamenávanie inflácie Českým štatistickým úradom. Zdrojmi tejto kapitoly sú publikácie [3], [11] a [12].

Druhá kapitola popisuje korelačnú analýzu, ktorá sa zaoberá overovaním závislostí medzi jednotlivými premennými s využitím koeficientov korelácie, mnohonásobnej korelácie a parciálnej korelácie. Tretia kapitola sa zaoberá regresnou analýzou, pričom je zameraná na model lineárnej regresie. Dôležitou súčasťou regresnej analýzy je overovanie presnosti modelu, na čo sa dá použiť koeficient determinácie (poprípade korigovaný koeficient determinácie). V prípade odhadovania parametrov metódou najmenších štvorcov je taktiež dôležité analyzovať možné problémy, akými sú autokorelovanosť rezíduí alebo multikolinearita. V štvrtej kapitole je popísaná analýza jednorozmerných časových rád pomocou Boxovej-Jenkinsovej metodológie. Z literatúry použitej na tieto tri kapitoly je možné spomenúť publikácie [1], [2] a [7].

Posledná kapitola sa zaoberá modelovaním miery inflácie v USA v programe Wolfram Mathematica 8 pomocou popísaných metód. V prvom príklade je aplikovaná korelačná a regresná analýza. Druhý príklad je zameraný na zvolenie vhodného ARMA modelu pomocou Boxovej-Jenkinsovej metodológie. Dáta použité v príkladoch sú získane zo zdroja [14].

# 1 Základné pojmy

## 1.1 Indexové čísla

*Indexové čísla* (Indexy) slúžia ako prostriedky na porovnávanie nejakých javov (agregátov) medzi miestami v jednom čase alebo na rovnakom mieste v rôznych časoch. Zaoberať sa budeme druhým prípadom, a teda porovnávaním javov v závislosti od času. Príkladom takýchto indexov je RPI (Retail Price Index, používaný v Spojenom Kráľovstve), CPI (Consumer Price Index, používaný v USA), PX (Prague Stock Exchange), FTSE (Financial Times Stock Exchange), S&P 500, Dow Jones a iné.

Nech  $A$  je nejaký cenový agregát (napríklad náklady na život), ktorý je rozdelený na zložky  $A_1, \dots, A_n$  (potraviny, bývanie, služby atď.). V čase  $t$  vyjadríme  $A_i$  ako súčin ceny  $p_t^i$  s váhou  $q_t^i$ ,  $t = 0, \dots, T$ ,  $i = 1, \dots, n$ . Označme  $\mathbf{p}_t := (p_t^1, \dots, p_t^n)^\top$  ako vektor cien a  $\mathbf{q}_t := (q_t^1, \dots, q_t^n)^\top$  vektor množstiev (váh). *Cenový Index* (Price Index) meria zmenu ceny komodít v priebehu času. V našom príklade by cenový index predstavoval číslo ukazujúce, ako sa agregát  $A$  mení so zmenami cien  $\mathbf{p}$ . Zvyčajne sa vyjadruje ako vážený priemer cien komodít v daných dvoch časových obdobiach.

Nech  $s$  je základné obdobie a  $t$  je bežné obdobie.  $I_{st}$  je cenový index, ktorý opisuje zmenu cenovej hladiny z času  $s$  do času  $t$ . Z historického hľadiska bol prvý index podielom dvoch aritmetických priemerov cien v obdobiach  $s$  a  $t$ :

$$I_{st} = \frac{\sum_{i=1}^n p_t^i}{\sum_{i=1}^n p_s^i} = \frac{\mathbf{1}^\top \cdot \mathbf{p}_t}{\mathbf{1}^\top \cdot \mathbf{p}_s}. \quad (1.1.1)$$

Iný, navrhnutý Edgeworthom, bol geometrický priemer cenových podielov:

$$I_{st}^E = \sqrt[n]{\prod_{i=1}^n \frac{p_t^i}{p_s^i}}. \quad (1.1.2)$$

V oboch príkladoch je problém, že index nepočíta s množstvami jednotlivých zložiek. Využiteľnejšie indexy, ktoré už zahŕňajú množstvá, sú napríklad Laspeyresov index a Paascheov index:

$$I_{st}^L = \frac{\sum_{i=1}^n p_t^i q_s^i}{\sum_{i=1}^n p_s^i q_s^i} = \frac{\mathbf{p}_t \cdot \mathbf{q}_s}{\mathbf{p}_s \cdot \mathbf{q}_s}, \quad (1.1.3)$$

$$I_{st}^P = \frac{\sum_{i=1}^n p_t^i q_t^i}{\sum_{i=1}^n p_s^i q_t^i} = \frac{\mathbf{p}_t \cdot \mathbf{q}_t}{\mathbf{p}_s \cdot \mathbf{q}_t}. \quad (1.1.4)$$

Geometrickým priemerom Laspeyresovho a Paascheovho indexu získame Fisherov index:

$$I_{st}^F = \sqrt{\left(\frac{\sum_{i=1}^n p_t^i q_s^i}{\sum_{i=1}^n p_s^i q_s^i}\right) \left(\frac{\sum_{i=1}^n p_t^i q_t^i}{\sum_{i=1}^n p_s^i q_t^i}\right)}. \quad (1.1.5)$$

## 1.2 Inflácia

*Inflácia* by sa dala definovať ako rast všeobecnej cenovej hladiny v priebehu času, čoho dôsledkom je zníženie kúpnej sily peňazí. Extrémnym prípadom inflácie je hyperinflácia, kedy ročná miera inflácie dosahuje až nad 1000%. Príkladom hyperinflácie je štát Zimbabwe, v ktorom za rok 2007 dosiahla ročná miera inflácie vyše 24000%. Opakom k inflácií (a aj zriedkavejším) je *deflácia*, teda pokles cenovej hladiny, ku ktorej došlo napríklad v Spojenom Kráľovstve v období 1920 až 1935.

Na meranie inflácie sa používa *miera inflácie*, ktorá sa počíta pomocou cenového indexu (budeme ho označovať RPI). Pokles miery inflácie sa nazýva *dezinflácia*. RPI zvyčajne zostavuje *štatistický úrad* nejakou modifikáciou Laspeyresovho cenového indexu, pričom váhy sú odvodené zo zloženia spotrebného koša. Jednotlivé váhy sa časom menia. Napríklad v Českej republike v roku 1999 bola váha potravín približne 198/1000, v roku 2005 približne 163/1000 a v roku 2007 približne 170/1000.

## 1.3 RPI v Českej republike

Všetky zložky spotrebného koša sú rozdelené do 12 základných skupín s celkovou váhou 1000. Skupiny a ich váhy sú od roku 2008 nasledovné:

1. Potraviny a nealkoholické nápoje (170,33)
2. Alkoholické nápoje, tabak (85,99)
3. Odev a obuv (47,21)
4. Bývanie, voda, energie a palivá (253,40)
5. Bytové vybavenie, zariadenie domácnosti, opravy (55,19)
6. Zdravie (25,00)
7. Doprava (114,89)
8. Pošty a telekomunikácie (39,88)
9. Rekreačia a kultúra (93,68)
10. Vzdelávanie (7,77)
11. Stravovanie a ubytovanie (44,26)
12. Ostatný tovar a služby (62,40)

Označme tieto váhy ako  $q_{2008}$ . Mieru mesačnej inflácie v novembri 2010 (označme ako 11/2010) získame z indexu

$$I_{10/2010, 11/2010} = \frac{p_{11/2010}^T \cdot q_{2008}}{p_{10/2010}^T \cdot q_{2008}} = 1,002. \quad (1.3.1)$$

Z čoho dostaneme mesačnú mieru inflácie 0,2%. Na porovnanie v októbri 2010 bola inflácia -0,2% (deflácia), a v decembri 2010 0,5%. Hodnoty môžeme vidieť v tab. 1.4.4.

Na výpočet ročnej miery inflácie pre obdobie od decembra 2009 do novembra 2010 (12/2009–11/2010) použijeme index

$$I_{12/2008-11/2009, 12/2009-11/2010} = \frac{p_{12/2009-11/2010}^T \cdot q_{2008}}{p_{12/2008-11/2009}^T \cdot q_{2008}} = 1,014. \quad (1.3.2)$$

A teda ročná miera inflácie pre dané obdobie (dĺžky jedného roku) je 1,4%, čo je hodnota z tab. 1.4.2.

## 1.4 Zaznamenávanie inflácie Českým štatistickým úradom

- 1) Miera inflácie (typ 1) je vyjadrená prírastkom priemerného ročného cenového indexu. Vyjadruje percentuálnu zmenu priemernej cenovej hladiny za posledných 12 mesiacov oproti priemeru 12-tich predchádzajúcich mesiacov. Táto miera inflácie sa používa pri prepočtoch reálnych miezd, dôchodkov a pod.

Rok										
00	01	02	03	04	05	06	07	08	09	10
3,9	4,7	1,8	0,1	2,8	1,9	2,5	2,8	6,3	1	1,5

Tab. 1.4.1. Miera inflácie (v %) pre kalendárne roky 2000-2010 (typ 1)

Rok	Mesiac											
	1	2	3	4	5	6	7	8	9	10	11	12
2000	2,1	2,2	2,3	2,4	2,5	2,6	2,9	3,1	3,3	3,6	3,8	3,9
2001	4	4	4	4,1	4,2	4,3	4,5	4,6	4,7	4,7	4,7	4,7
2002	4,6	4,6	4,6	4,5	4,3	3,9	3,5	3,1	2,7	2,4	2,1	1,8
2003	1,5	1,1	0,8	0,5	0,3	0,2	0,2	0,1	0	0	0,1	0,1
2004	0,3	0,5	0,8	1	1,2	1,4	1,7	2	2,2	2,5	2,7	2,8
2005	2,8	2,7	2,6	2,6	2,5	2,4	2,2	2,1	2	2	1,9	1,9
2006	2	2,1	2,2	2,3	2,4	2,5	2,6	2,7	2,8	2,7	2,6	2,5
2007	2,4	2,3	2,2	2,2	2,1	2,1	2,1	2	2	2,2	2,5	2,8
2008	3,4	3,9	4,3	4,7	5	5,4	5,8	6,1	6,4	6,6	6,5	6,3
2009	5,9	5,4	5	4,6	4,1	3,7	3,1	2,6	2,1	1,6	1,3	1
2010	0,9	0,8	0,7	0,6	0,6	0,6	0,8	0,9	1,1	1,2	1,4	1,5

Tab. 1.4.2. Miera inflácie (v %) pre dané obdobia dĺžky 12 mesiacov (typ 1)



- 2) Miera inflácie (typ 2) vyjadrená prírastkom cenového indexu k rovnakému mesiacu predchádzajúceho roku. Vyjadruje percentuálnu zmenu cenovej hladiny vo vykazovanom mesiaci daného roku. Používa sa pri prepočtoch reálnej úrokovej miery, reálneho zvýšenia cien majetku, valorizácií a pod.

Rok	Mesiac											
	1	2	3	4	5	6	7	8	9	10	11	12
1997	7,4	7,3	6,8	6,7	6,3	6,8	9,4	9,9	10,3	10,2	10,1	10
1998	13,1	13,4	13,4	13,1	13	12	10,4	9,4	8,8	8,2	7,5	6,8
1999	3,5	2,8	2,5	2,5	2,4	2,2	1,1	1,4	1,2	1,4	1,9	2,5
2000	3,4	3,7	3,8	3,4	3,7	4,1	3,9	4,1	4,1	4,4	4,3	4
2001	4,2	4	4,1	4,6	5	5,5	5,9	5,5	4,7	4,4	4,2	4,1
2002	3,7	3,9	3,7	3,2	2,5	1,2	0,6	0,6	0,8	0,6	0,5	0,6
2003	-0,4	-0,4	-0,4	-0,1	0	0,3	-0,1	-0,1	0	0,4	1	1
2004	2,3	2,3	2,5	2,3	2,7	2,9	3,2	3,4	3	3,5	2,9	2,8
2005	1,7	1,7	1,5	1,6	1,3	1,8	1,7	1,7	2,2	2,6	2,4	2,2
2006	2,9	2,8	2,8	2,8	3,1	2,8	2,9	3,1	2,7	1,3	1,5	1,7
2007	1,3	1,5	1,9	2,5	2,4	2,5	2,3	2,4	2,8	4	5	5,4
2008	7,5	7,5	7,1	6,8	6,8	6,7	6,9	6,5	6,6	6	4,4	3,6
2009	2,2	2	2,3	1,8	1,3	1,2	0,3	0,2	0	-0,2	0,5	1
2010	0,7	0,6	0,7	1,1	1,2	1,2	1,9	1,9	2	2	2	2,3

Tab. 1.4.3. Miera inflácie (v %) pre dané obdobie dĺžky 12 mesiacov (typ 2)

- 3) Miera inflácie (typ 3) vyjadrená prírastkom cenového indexu k predchádzajúcemu mesiacu. Vyjadruje percentuálnu zmenu cenovej hladiny daného mesiaca oproti predchádzajúcemu mesiacu.

Rok	Mesiac											
	1	2	3	4	5	6	7	8	9	10	11	12
1997	1,2	0,3	0,1	0,6	0,1	1,2	3,5	0,7	0,6	0,4	0,4	0,5
1998	4	0,6	0,1	0,3	0,1	0,3	1,9	-0,2	0,1	-0,2	-0,2	-0,2
1999	0,8	0	-0,2	0,3	-0,1	0,2	0,8	0,1	-0,1	0	0,2	0,5
2000	1,7	0,2	0	-0,1	0,2	0,6	0,6	0,2	0	0,3	0,1	0,2
2001	1,9	0	0,1	0,4	0,6	1	1	-0,2	-0,7	0	-0,1	0,1
2002	1,5	0,2	-0,1	-0,1	-0,1	-0,3	0,5	-0,2	-0,5	-0,3	-0,2	0,2
2003	0,6	0,2	-0,1	0,2	0	0	0,1	-0,2	-0,5	0,1	0,5	0,2
2004	1,8	0,2	0,1	0	0,4	0,2	0,4	0	-0,8	0,5	-0,1	0,1
2005	0,7	0,2	-0,1	0,1	0,2	0,6	0,3	0	-0,3	0,9	-0,3	-0,1
2006	1,4	0,1	-0,1	0,1	0,5	0,3	0,4	0,2	-0,7	-0,5	-0,1	0,2
2007	1	0,3	0,3	0,7	0,4	0,3	0,4	0,3	-0,3	0,6	0,9	0,5
2008	3	0,3	-0,1	0,4	0,5	0,2	0,5	-0,1	-0,2	0	-0,5	-0,3
2009	1,5	0,1	0,2	-0,1	0	0	-0,4	-0,2	-0,4	-0,2	0,2	0,2
2010	1,2	0	0,3	0,3	0,1	0	0,3	-0,3	-0,3	-0,2	0,2	0,5

Tab. 1.4.4. Miera inflácie (v %) pre dané obdobie dĺžky 1 mesiac (typ 3)

## 2 Korelačná analýza

### 2.1 Korelačný koeficient

Nech  $X$  a  $Y$  sú náhodné veličiny s konečnými druhými momentmi a s kladnými rozptylmi. Potom definujeme *korelačný koeficient* ako

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}X \text{var}Y}}. \quad (2.1.1)$$

Niekedy používame značenie  $\text{cor}(X, Y)$  alebo tiež  $\rho(X, Y)$ .

Vlastnosti korelačného koeficientu:

- 1) Pre korelačný koeficient platí  $-1 \leq \rho_{XY} \leq 1$ . Rovnosť  $\rho_{XY} = 1$  platí práve vtedy, keď  $Y = a + bX$ , pričom  $b > 0$ ,  $\rho_{XY} = -1$  platí práve vtedy, keď  $Y = a + bX$ , pričom  $b < 0$ .
- 2) Korelácia meria silu lineárnej závislosti dvoch náhodných veličín. Pre  $|\rho_{XY}| = 1$  je závislosť lineárna. Ak je  $\rho_{XY} = 0$ , tak hovoríme, že veličiny  $X, Y$  sú *nekorelované*. To znamená, že závislosť medzi nimi nie je možné modelovať priamkou.
- 3) Pre ľubovoľné konštanty  $a, b, c, d$  také, že  $bd \neq 0$  platí

$$\rho_{a+bX, c+dY} = \text{sgn}(bd) \rho_{XY}. \quad (2.1.2)$$

Majme náhodne vektory  $\mathbf{X} = (X_1, \dots, X_n)^T$  a  $\mathbf{Y} = (Y_1, \dots, Y_m)^T$  s konečnými druhými momentmi. Nech ďalej platí  $\text{var}X_i > 0$ ,  $i = 1, \dots, n$  a  $\text{var}Y_j > 0$ ,  $j = 1, \dots, m$ . *Korelačnú maticu*  $\text{cor}(\mathbf{X}, \mathbf{Y})$  rozumieme maticu typu  $n \times m$ , ktorej  $(i, j)$ -tý prvok je  $\rho(X_i, Y_j)$ . Špeciálnym prípadom je matica  $\text{cor}(\mathbf{X}, \mathbf{X}) = \text{cor}\mathbf{X}$  s prvkami  $\rho(X_i, X_j)$ ,  $i, j = 1, \dots, n$ .

Majme súbor dvojíc pozorovaní  $(x_1, y_1), \dots, (x_n, y_n)$  náhodných veličín  $X$  a  $Y$ . Potom definujeme *výberovú kovarianciu* ako

$$S_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}), \quad (2.1.3)$$

kde  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  a  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ .

Špeciálne dostávame *výberový rozptyl*

$$S_{xx} = S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (2.1.4)$$

Ak je  $S_x^2 > 0$  a  $S_y^2 > 0$ , tak definujeme výberový korelačný koeficient  $r_{xy}$  vzorcom

$$r_{xy} = \frac{S_{xy}}{S_x S_y}. \quad (2.1.5)$$

Majme  $p$  pozorovaní náhodného vektora  $\mathbf{X} = (X_1, \dots, X_n)^\top$ , tvoriace riadky matice

$$\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \dots & x_{pn} \end{pmatrix},$$

$i$ -tý riadok označme  $\mathbf{x}_i^\top = (x_{i1}, \dots, x_{in})$ ,  $i = 1, \dots, p$ . Výberová variančná matica náhodného vektora  $\mathbf{X}$  je

$$\mathbf{S} = \frac{1}{p-1} \sum_{i=1}^p (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top, \text{ kde } \bar{\mathbf{x}} = \frac{1}{p} \sum_{i=1}^p \mathbf{x}_i. \quad (2.1.6)$$

Označme  $s_{ij}$  prvky matice  $\mathbf{S}$ . Výberový korelačný koeficient píšeme v tvare

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}}. \quad (2.1.7)$$

Výberová korelačná matica  $\mathbf{R}$  má prvky  $r_{ij}$ , kde  $i, j = 1, \dots, n$ .

## 2.2 Koeficient mnohonásobnej korelácie

Majme náhodnú veličinu  $Y$  a náhodný vektor  $\mathbf{X} = (X_1, \dots, X_n)^\top$  s konečnými druhými momentmi. Nech  $\text{var}Y > 0$  a nech  $\mathbf{V} = \text{var}\mathbf{X}$  je regulárna matica. Závislosť medzi  $Y$  a vektorom  $\mathbf{X}$  meriame pomocou koeficientu mnohonásobnej korelácie, ktorý je definovaný ako

$$\rho_{Y,\mathbf{X}}^{(M)} = \rho_{Y, \alpha + \beta^\top \mathbf{X}}, \quad (2.2.1)$$

čo je korelačný koeficient medzi veličinou  $Y$  a jej najlepšou lineárnou aproximáciou  $\hat{Y} = \alpha + \beta^\top \mathbf{X}$ .

Vlastnosti koeficientu mnohonásobnej korelácie:

- 1) Na určenie koeficientov  $\alpha$  a  $\beta$  minimalizujeme hodnotu  $E(Y - \alpha - \beta^\top \mathbf{X})^2$ . Dá sa ukázať, že minimum nastáva pre koeficienty  $\alpha = EY - \beta^\top E\mathbf{X}$  a  $\beta = \mathbf{V}^{-1} \text{cov}(\mathbf{X}, Y)$ , kde  $\text{cov}(\mathbf{X}, Y)$  je stĺpcový vektor so zložkami  $\text{cov}(X_1, Y), \dots, \text{cov}(X_n, Y)$ .
- 2) Koeficient mnohonásobnej korelácie vyjadruje silu lineárnej závislosti medzi náhodnou veličinou  $Y$  a náhodným vektorom  $\mathbf{X}$ .

- 3) Koeficient mnohonásobnej korelácie je vždy nezáporný a nie je nikdy menší ako absolútna hodnota ktoréhokoľvek obyčajného korelačného koeficientu  $\rho(X_i, Y)$ ,  $i = 1, \dots, n$ .
- 4) Koeficient mnohonásobnej korelácie môžeme písať v tvare

$$\rho_{Y,X}^{(M)} = \sqrt{\text{cor}(Y, \mathbf{X})(\text{cor} \mathbf{X})^{-1} \text{cor}(\mathbf{X}, Y)}, \quad (2.2.2)$$

kde  $\text{cor}(Y, \mathbf{X})$  je riadkový vektor so zložkami  $\rho(X_1, Y), \dots, \rho(X_n, Y)$  a  $\text{cor}(\mathbf{X}, Y)$  je jeho transpozícia.

Majme  $p$  pozorovaní veličiny  $Y$  a vektora  $\mathbf{X}$ . Odhadneme po zložkách vektor  $\text{cor}(\mathbf{X}, Y)$  a maticu  $\text{cor} \mathbf{X}$  a označíme príslušné odhady ako  $R_{\mathbf{X},Y}$  a  $R$ . Potom dostaneme výberový koeficient mnohonásobnej korelácie ako

$$r_{Y\mathbf{X}}^{(M)} = \sqrt{R_{Y\mathbf{X}} R^{-1} R_{\mathbf{X}Y}}, \quad (2.2.3)$$

kde  $R_{Y\mathbf{X}} = R_{\mathbf{X}Y}^T$ .

## 2.3 Koeficient parciálnej korelácie

Hodnota korelácie dvoch náhodných veličín môže byť ovplyvnená pôsobením ďalších veličín na skúmanú dvojicu. Preto zavádzame koeficient parciálnej korelácie.

Majme dve náhodné veličiny  $Y$  a  $Z$ , ktoré môžu byť ovplyvňované veličinami  $X_1, \dots, X_n$ . Predpokladajme, že veličiny  $Y$  a  $Z$  majú konečné druhé momenty a že  $\mathbf{X} = (X_1, \dots, X_n)^T$  má regulárnu variančnú maticu  $\mathbf{V}$ . Zaujímá nás „čistá“ korelácia medzi  $Y$  a  $Z$ , ak sa vplyv vektora  $\mathbf{X}$  eliminuje. Táto eliminácia by mohla spočívať v tom, že by sme veličiny  $Y$  a  $Z$  sledovali za takých podmienok, pri ktorých sa vektor  $\mathbf{X}$  nemení. Ak sa obmedzíme na lineárne vzťahy, tak najlepšou aproximáciou veličiny  $Y$  pomocou vektora  $\mathbf{X}$  je  $\hat{Y} = \alpha_1 + \beta_1^T \mathbf{X}$ , kde

$$\alpha_1 = EY - \beta_1^T EX \text{ a } \beta_1 = \mathbf{V}^{-1} \text{cov}(\mathbf{X}, Y). \quad (2.3.1)$$

Tú časť veličiny, ktorú vektor  $\mathbf{X}$  nevysvetlí, si predstavujeme ako rezíduum  $Y - \hat{Y}$ . Podobne aproximujeme  $Z$  pomocou  $\hat{Z} = \alpha_2 + \beta_2^T \mathbf{X}$ , kde

$$\alpha_2 = EZ - \beta_2^T EX \text{ a } \beta_2 = \mathbf{V}^{-1} \text{cov}(\mathbf{X}, Z). \quad (2.3.2)$$

Za predpokladu  $\text{var}(Y - \hat{Y}) > 0$  a  $\text{var}(Z - \hat{Z}) > 0$  definujeme *parciálny korelačný koeficient*  $\rho_{YZ|\mathbf{X}}^{(P)}$  veličín  $Y$  a  $Z$  pri pevnom  $\mathbf{X}$  ako korelačný koeficient  $\rho_{Y-\hat{Y}, Z-\hat{Z}}$ .

Vlastnosti koeficientu parciálnej korelácie:

- 1) Pretože na hodnotách  $\alpha_1$  a  $\alpha_2$  koeficient  $\rho_{Y-\hat{Y}, Z-\hat{Z}}$  nezávisí, platí

$$\rho_{YZ|\mathbf{X}}^{(P)} = \rho_{Y-\beta_1^T \mathbf{X}, Z-\beta_2^T \mathbf{X}}. \quad (2.3.3)$$

2) Ak je koeficient  $\rho_{YZ|X}^{(P)}$  definovaný, tak platí

$$\rho_{YZ|X}^{(P)} = \frac{\rho_{YZ} - \text{cor}(Y, \mathbf{X})(\text{cor} \mathbf{X})^{-1} \text{cor}(\mathbf{X}, Z)}{\sqrt{(1 - (\rho_{YX}^{(M)})^2)(1 - (\rho_{ZX}^{(M)})^2)}}. \quad (2.3.4)$$

Ak dosadíme do (2.3.4) výberové hodnoty, dostaneme *výberový parciálny koeficient*

$$r_{YZ|X}^{(P)} = \frac{r_{YZ} - R_{YX}R^{-1}R_{XZ}}{\sqrt{(1 - (r_{YX}^{(M)})^2)(1 - (r_{ZX}^{(M)})^2)}}. \quad (2.3.5)$$

V praxi sa väčšinou stretávame s prípadom, že  $r_{YZ|X}^{(P)} < r_{YZ}$ , avšak môže nastať aj opačná nerovnosť (viď [4]). Napríklad v prípade troch veličín  $X, Y$  a  $Z$  je výberový parciálny korelačný koeficient veličín  $X, Y$  pri pevnom  $Z$  definovaný vzorcom

$$r_{YX|Z}^{(P)} = \frac{r_{YX} - r_{YZ}r_{XZ}}{\sqrt{(1 - r_{YZ}^2)(1 - r_{XZ}^2)}}. \quad (2.3.6)$$

Potom ak  $r_{YZ}$  a  $r_{XZ}$  majú rovnaké znamienka tak  $r_{YX} - r_{YZ}r_{XZ} < r_{YX}$ . V opačnom prípade, ak  $r_{YZ}$  a  $r_{XZ}$  majú rozdielne znamienka, tak  $r_{YX} - r_{YZ}r_{XZ} > r_{YX}$ . A teda môže nastať  $r_{YX|Z}^{(P)} > r_{YX}$  tým skôr, čím bližšie je  $r_{YZ}$  a  $r_{XZ}$  k nule. Nasledujúci postup výpočtu parciálnych korelácií je prevzatý z knihy [6].

Označme  $\mathbf{D} = (\text{cor} \mathbf{X})^{-1} = (d_{i,j})_{i,j=1}^n$ . Definujme *škálovaciu maticu*  $\mathbf{S}_D$  vzorcom

$$\mathbf{S}_D = \text{diag} \left\{ \frac{1}{\sqrt{d_{11}}}, \dots, \frac{1}{\sqrt{d_{nn}}} \right\}. \quad (2.3.7)$$

Potom matica

$$\mathbf{P} = (-1)\mathbf{S}_D\mathbf{D}\mathbf{S}_D \quad (2.3.8)$$

je matica parciálnych korelácií vektora  $\mathbf{X} = (X_1, \dots, X_n)^T$ , kde na diagonále sú hodnoty (-1) a mimo diagonálu sú parciálne korelácie  $\rho_{X_i X_j}^{(P)}$  pri pevných  $X_k$ , kde  $k \neq i, j$ .

Ak označíme  $\hat{\mathbf{D}} = \mathbf{R}^{-1}$ , tak dostaneme maticu *výberových parciálnych korelácií* vzorcom

$$\hat{\mathbf{P}} = (-1)\mathbf{S}_{\hat{\mathbf{D}}}\hat{\mathbf{D}}\mathbf{S}_{\hat{\mathbf{D}}}. \quad (2.3.9)$$

### 3 Regresná analýza

Regresná analýza je jeden z najdôležitejších nástrojov, slúžiacich na popis náhodnej veličiny  $Y$  (regresand alebo tiež závisle premenná) pomocou veličín  $X_1, \dots, X_k$  (regresory alebo nezávisle premenné). Inak povedané, úlohou regresie je vysvetliť zmeny hodnôt jednej premennej zmenami hodnôt iných premenných. To znamená, že hľadáme funkciu  $f$  takú, že  $Y = f(\mathbf{X})$ ,  $\mathbf{X} = (X_1, \dots, X_k)^\top$ .

#### 3.1 Lineárna regresia

Vzťah medzi závisle premennou  $Y$  a vektorom regresorov vyjadríme pomocou lineárnej funkcie. Závislosť modelujeme rovnicou

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + e, \quad (3.1.1)$$

kde  $e$  je náhodná chyba splňujúca podmienky  $Ee = 0$ ,  $\text{var } e = \sigma^2$ , kde  $\sigma^2 > 0$  je neznámy parameter.

Nech  $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$  je vektor  $n$  realizácií závisle premennej a  $\mathbf{X}$  je pevná matica typu  $n \times (k + 1)$ , ktorá obsahuje v nultom stĺpci jednotky a v  $j$ -tom stĺpci  $n$  pozorovaní veličiny  $X_j$ ,  $j = 1, \dots, k$ .

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix}. \quad (3.1.2)$$

Pre jednotlivé realizácie máme vzťah

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + e_i, \quad i = 1, \dots, n, \quad (3.1.3)$$

kde  $\mathbf{e} = (e_1, \dots, e_n)^\top$  je vektor náhodných chýb (rezíduí), ktorý splňuje  $E\mathbf{e} = \mathbf{0}$ ,  $\text{var } \mathbf{e} = \sigma^2 \mathbf{I}_n$ , kde  $\sigma^2 > 0$  a  $\mathbf{I}_n$  je jednotková matica.

Vektor parametrov  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_k)^\top$  odhadujeme *metódou najmenších štvorcov* (*Ordinary Least Squares*, OLS). To znamená, že minimalizujeme súčet štvorcov  $S$  vzhľadom k  $\boldsymbol{\beta}$ , kde

$$S = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij})^2. \quad (3.1.4)$$

Riešením je vektor odhadov  $\mathbf{b} = (b_0, \dots, b_k)^\top$ . Vektor  $\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b}$  je potom lineárna aproximácia vektora  $\mathbf{Y}$  odhadnutým regresným modelom. Veličina

$$S_e = (\mathbf{Y} - \hat{\mathbf{Y}})^\top (\mathbf{Y} - \hat{\mathbf{Y}}) = \mathbf{Y}^\top \mathbf{Y} - \mathbf{b}^\top \mathbf{X}^\top \mathbf{Y} \quad (3.1.5)$$

sa nazýva *reziduálny súčet štvorcov*.

Vlastnosti odhadu  $\mathbf{b}$ :

- 1) Vektor odhadov  $\mathbf{b}$  je možné vyjadriť ako

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (3.1.6)$$

- 2) Odhad  $\mathbf{b}$  metódou najmenších štvorcov je nestranný.

Nestranný odhad parametra  $\sigma^2$  je

$$s^2 = \frac{S_e}{(n - k - 1)}. \quad (3.1.7)$$

### 3.2 Koeficient determinácie

Na zistenie presnosti modelu (zhoda s dátami) sa používa *koeficient determinácie*  $R^2$ , ktorý predstavuje variabilitu veličiny  $Y$ , ktorú vysvetlí regresný model pri daných dátach  $y_1, \dots, y_n$ . Je definovaný predpisom

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (3.2.1)$$

kde  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  a  $\hat{y}_i = b_0 + \sum_{j=1}^k b_j x_{ij}$ .

Koeficient determinácie je kvadrát výberového koeficientu mnohonásobnej korelácie medzi  $Y$  a vektorom nezávislých premenných  $\mathbf{X}$ . Teda platí  $0 \leq R^2 \leq 1$ . Ak je táto korelácia vysoká (blízko 1), znamená to, že sa model s danými dátami zhoduje dobre. Koeficient  $R^2$  sa používa pre posúdenie významnosti vplyvu jednotlivých regresorov. V prípade, že vylúčenie niektorej nezávisle premennej spôsobí len nepatrné zníženie koeficientu determinácie, znamená to, že príslušný regresor má zanedbateľný prínos pre popis variability veličiny  $Y$ , a teda model môžeme zjednodušiť. Avšak v praxi sa môže stať, že model s vysokým  $R^2$  má tendenciu byť príliš komplikovaný. V takom prípade dávame prednosť korigovanému koeficientu determinácie  $\bar{R}^2$ , ktorý penalizuje nadmerný počet regresorov. Je definovaný ako

$$\bar{R}^2 = 1 - \left[ \frac{n-1}{n-k-1} (1 - R^2) \right]. \quad (3.2.2)$$

### 3.3 Normálny model lineárnej regresie

O normálnom modeli hovoríme v prípade, keď predpokladáme  $\mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ . Z toho vyplýva, že vektor  $\mathbf{Y}$  pochádza z normálneho rozdelenia  $N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$ . A potom platí aj  $\mathbf{b} \sim N(\boldsymbol{\beta}, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1})$ .

### 3.4 Testy nulovosti parametrov

Ak nás zaujíma, ktoré regresory skutočne do modelu patria, a teda významne ovplyvňujú veličinu  $Y$ , môžeme použiť metódy korelačnej analýzy, alebo v rámci regresnej analýzy testy nulovosti parametrov  $\beta_1, \dots, \beta_k$ . Predpokladmi testu nulovosti sú lineárna nezávislosť stĺpcov matice  $X$  (predpoklad už pri konštrukcii odhadu OLS pre  $\beta$ ) a normalita modelu.

Potom testujeme nulovú hypotézu

$$H_0 : \beta_j = 0, j = 0, 1, \dots, k \quad (3.4.1)$$

proti alternatíve

$$H_1 : \beta_j \neq 0, j = 0, 1, \dots, k. \quad (3.4.2)$$

Označme  $v_{jj}$  diagonálny prvok matice  $(X^T X)^{-1}$ . Kritický obor testu píšeme v tvare

$$H_0 \text{ zamietame na hladine } \alpha \Leftrightarrow \frac{|b_j|}{s\sqrt{v_{jj}}} \geq t_{n-k-1} \left(1 - \frac{\alpha}{2}\right), \quad (3.4.3)$$

kde  $t_{n-k-1} \left(1 - \frac{\alpha}{2}\right)$  je  $\left(1 - \frac{\alpha}{2}\right)$ -kvantil Študentovho rozdelenia. Ak nulovú hypotézu pre  $j \geq 1$  nezamietame, znamená to, že  $Y$  nezávisí v uvažovanom modeli na  $X_j$ .

### 3.5 Autokorelovanosť rezíduí

O *autokorelovanosti rezíduí* hovoríme v prípade, kedy reziduálna zložka  $e_i$  je korelovaná so svojimi oneskorenými a budúcimi hodnotami. Korelovanosť je pomerne obvyklá pre veličiny s časovým usporiadaním, čo znamená, že sa s ňou môžeme stretnúť aj pri modelovaní inflácie.

Príčiny autokorelovanosti sú napríklad:

- 1) Nepoužitie niektorých regresorov, ktorých časové rady vykazujú autokorelovanosť, a tá sa presunie do reziduálnej zložky.
- 2) Pozorovania  $y_s$  a  $y_t$  majú viac spoločného ako daný model.
- 3) Nezaradenie oneskorených hodnôt regresandu alebo niektorých regresorov.

Najjednoduchší typ autokorelovanosti spočíva v modelovaní reziduálnej zložky  $e_i$  pomocou *autoregresného modelu* prvého rádu (AR(1)), ktorý je v tvare

$$e_i = \varphi e_{i-1} + u_i, \quad (3.5.1)$$

kde  $-1 < \varphi < 1$  je parameter a  $u_i$  je *biely šum*, čo je označenie časovej rady navzájom nekorelovaných veličín s nulovou strednou hodnotou a konštantným kladným rozptylom. Kladné  $\varphi$  (*pozitívna autokorelovanosť*) znamená zotrvačnosť



v znamienku susedných hodnôt  $e_i$ . Naopak záporné  $\varphi$  (*negatívna autokorelovanosť*) hovorí o častejších zmenách v znamienku susedných hodnôt  $e_i$ .

V modeli AR(1) pre ľubovoľné  $i$  platí

$$\varphi = \rho = \text{cor}(e_i, e_{i-1}) \quad (3.5.2)$$

a môžeme použiť *Durbinov-Watsonov test* autokorelovanosti rezíduí s nulovou hypotézou  $H_0 : \rho = 0$ . Testová štatistika má tvar

$$DW = \frac{\sum_{i=2}^n (\hat{e}_i - \hat{e}_{i-1})^2}{\sum_{i=1}^n \hat{e}_i^2}, \text{ kde } \hat{e}_i = y_i - \hat{y}_i. \quad (3.5.3)$$

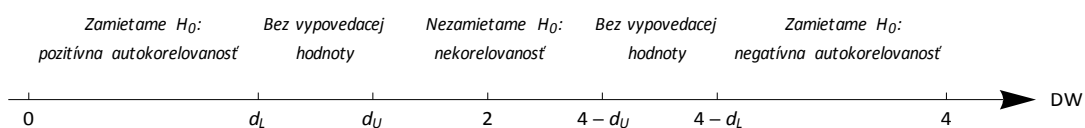
Dá sa odvodiť aproximácia

$$DW \approx 2(1 - \hat{\varphi}), \quad (3.5.4)$$

kde  $\hat{\varphi}$  je OLS-odhad parametra  $\varphi$  v modeli (3.5.1). Táto aproximácia dáva tri dôležité hodnoty:

- 1) Ak  $\hat{\varphi} = 0 \Rightarrow DW \approx 2$ , tak susedné rezídua sú nekorelované.
- 2) Ak  $\hat{\varphi} = 1 \Rightarrow DW \approx 0$ , tak susedné rezídua sú extrémne pozitívne korelované.
- 3) Ak  $\hat{\varphi} = -1 \Rightarrow DW \approx 4$ , tak susedné rezídua sú extrémne negatívne korelované.

Za predpokladu normality bieleho šumu  $u_i$  má štatistika  $DW$  dve kritické hodnoty  $d_L$  (dolná) a  $d_U$  (horná), ktoré závisia od počtu pozorovaní  $n$ , počtu regresorov  $k$  a hladiny testu. Kritické hodnoty  $d_L$  a  $d_U$  môžeme nájsť v štatistických tabuľkách (viď napríklad [10]). Závěry vidíme na obr. 3.5.1.



**Obr. 3.5.1.** Závěry Durbinovho-Watsonovho testu pre hodnoty testovej štatistiky  $DW$

V praxi sa používajú zjednodušené pravidlá, kedy  $DW$  pod hodnotu 1,5 pri 50 a viac pozorovaniach a menšom počte regresorov znamená pozitívnu autokorelovanosť.

Ignorovanie autokorelovanosti rezíduí spôsobuje, že OLS-odhad  $\mathbf{b}$  nie je obecné najlepším nestranným odhadom parametra  $\boldsymbol{\beta}$ . Ďalej napríklad v prípade pozitívnej autokorelovanosti rezíduí odhady smerodajných odchýlok zložiek vektora  $\mathbf{b}$  bývajú podhodnotené, čo môže viesť v danom modeli k nesprávnemu prezentovaniu niektorých zložiek vektora parametrov  $\boldsymbol{\beta}$  ako významných (viď menovateľ v kritickom

obore (3.4.3)). A podobne tiež odhad smerodajnej odchýlky reziduálnej zložky  $\left(\frac{1}{n}\sum_{i=1}^n(y_i - \hat{y}_i)^2\right)^{\frac{1}{2}}$  býva podhodnotený, čo spôsobuje zvýšenie koeficientu determinácie (viď (3.2.1)).

Obecnejšie testy autokorelovanosti rezíduí sú popísané v knihe [2].

### 3.6 Multikolinearita

V praxi sa často stretávame s problémom *multikolinearity*, spočívajúcej v takmer lineárnej závislosti stĺpcov matice  $\mathbf{X}$ . To znamená, že nájdenie inverzie matice  $\mathbf{X}^T\mathbf{X}$  je numericky náročné (hovoríme o *zle podmienenej matici*). Zdroj [9] uvádza definíciu *podmienenosti matice A* ako

$$C_p = \|\mathbf{A}\| \cdot \|\mathbf{A}^{-1}\|, \quad (3.6.1)$$

kde  $\|\mathbf{A}\|$  je norma matice  $\mathbf{A}$ . O zle podmienenej matici hovoríme ak  $C_p \gg 1$ .

Zo štatistického hľadiska môžeme multikolinearitu interpretovať ako vysokú vzájomnú koreláciu medzi regresormi. Vo finančnej praxi je multikolinearita častá, pretože veličiny používané ako regresory majú často veľmi podobný vývojový trend. Dôsledkom multikolinearity je, že odhadnuté parametre príslušného regresného modelu majú vysoké rozptyly a môžu byť nepoužiteľné. Rozptyl odhadu parametra  $\beta_i$  je v tvare

$$\text{var}(b_i) = \frac{\sigma^2}{(n-1)s_{X_i}^2(1-R_i^2)}, \quad (3.6.2)$$

kde  $s_{X_i}^2$  je výberový rozptyl regresora  $X_i$

$$s_{X_i}^2 = \frac{1}{(n-1)} \sum_{t=1}^n (x_{ti} - \bar{x}_i)^2, \quad \bar{x}_i = \frac{1}{n} \sum_{t=1}^n x_{ti} \quad (3.6.3)$$

a  $R_i^2$  je koeficient determinácie pri lineárnej regresii premennej  $X_i$  na zostávajúce regresory pôvodného modelu. Definujme *faktor variančnej inflácie (variance inflation factor)* ako

$$VIF_i = \frac{1}{1-R_i^2}, \quad (3.6.4)$$

ktorý ukazuje koľkokrát sa zväčšil  $\text{var}(b_i)$  zásluhou multikolinearity. A teda pri menšom faktore variančnej inflácie získame presnejší odhad priameho vplyvu regresora  $X_i$  na veličinu  $Y$ .

Existujú jednoduché orientačné kritéria na rozpoznanie multikolinearity, kedy sa posudzujú nediagonálne prvky výberovej korelačnej matice  $\mathbf{R}$  alebo veľkosť výbero-

vého koeficientu mnohonásobnej korelácie. Ak je absolútna hodnota výberového korelačného koeficientu dvojice regresorov vysoká (napríklad nad 0,9), potom dvojica regresorov môže spôsobovať multikolinearitu a môžeme skúsiť jeden z nich vynechať. Alebo ak je vysoký koeficient mnohonásobnej korelácie medzi regresorom  $X_i$  a zostávajúcimi regresormi, potom zrejme regresor  $X_i$  spôsobuje multikolinearitu a môžeme ho skúsiť vynechať. Okrem toho existujú aj štatistické testy multikolinearity, ktoré sú založené na výberovom koeficiente mnohonásobnej korelácie alebo na výberovom parciálnom korelačnom koeficiente. Je ich možné nájsť v knihe [2].

Rôzne prístupy k modelom s multikolinearitou:

- 1) Ignorovanie multikolinearity (niekedy je model aj v takom prípade adekvátny, čo sa týka presnosti a schopnosti predpovedania).
- 2) Vynechanie regresorov, ktoré ju spôsobujú.
- 3) Transformácia regresorov (napríklad normovaním, centrovaním, prechodom k diferenciám atď.).
- 4) Rozšírenie dát (napríklad zvýšenie frekvencie pozorovaní).
- 5) Použitie metódy hlavných komponentov (viď [7]).

## 4 Analýza časových rád

V tejto kapitole sa budeme zaoberať analýzou časových rád pomocou *Boxovej-Jenkinsovej metodológie*, ktorá sa vyznačuje veľkým dôrazom na korelačnú analýzu.

### 4.1 Autokorelačná a parciálna autokorelačná funkcia

Definujme *stacionaritu* časovej rady  $\{y_i\}$  tak, že pre každé  $i$  a  $j$  platí

$$E(y_i) = \mu, \quad (4.1.1)$$

$$\text{cov}(y_i, y_j) = E(y_i - \mu)(y_j - \mu) = \text{cov}(y_{i+k}, y_{j+k}) \quad (4.1.2)$$

pre ľubovoľné  $k$ , a teda špeciálne aj

$$\text{var}(y_i) = \sigma_y^2. \quad (4.1.3)$$

To znamená, že stredná hodnota a rozptyl stacionárnej rady sú konštantné v čase.

Nech ďalej  $\{y_i\}$  je stacionárna časová rada s nulovou strednou hodnotou. Potom definujeme *autokovariančnú funkciu* pre oneskorenie  $k$  ako

$$\gamma_k = \text{cov}(y_i, y_{i-k}) = E(y_i - \mu)(y_{i-k} - \mu), \quad k = \dots, -1, 0, 1, \dots \quad (4.1.4)$$

Analogicky definujeme *autokorelačnú funkciu* (ACF) pre oneskorenie  $k$  ako

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_k}{\sigma_y^2} = \frac{\text{cov}(y_i, y_{i-k})}{\sqrt{\text{var}(y_i)}\sqrt{\text{var}(y_{i-k})}}, \quad k = \dots, -1, 0, 1, \dots \quad (4.1.5)$$

Autokovariančná a autokorelačná funkcia sa používa ku kvantitatívnemu popisu silnej korelovanosti časových rád v čase. Zrejme platí  $\gamma_k = \gamma_{-k}$  a  $\rho_k = \rho_{-k}$ , a teda sú to párne funkcie. Z toho dôvodu sa na ich popis stačí obmedziť na  $k \geq 0$ . Pritom platí  $\rho_0 = 1$  a  $|\rho_k| \leq 1$ . Na popis krátkodobej dynamiky stacionárnej rady slúži grafický záznam  $\rho_k$  pre jednotlivé  $k$ , ktorý sa nazýva *korelogram*.

Pre danú stacionárnu časovú radu definujeme *odhad autokovariančnej funkcie* ako

$$c_k = \frac{1}{n} \sum_{i=k+1}^n (y_i - \bar{y})(y_{i-k} - \bar{y}), \quad k = 0, 1, \dots, n-1, \quad (4.1.6)$$

kde  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ . *Odhad autokorelačnej funkcie* definujeme vzorcom

$$r_k = \frac{c_k}{c_0}, \quad k = 0, 1, \dots, n-1. \quad (4.1.7)$$

Chovanie ACF napovedá, aký model je v rámci Boxovej-Jenkinsovej metodológie vhodné použiť. Pre identifikáciu je predovšetkým dôležité určiť hodnotu  $k_0$  (bod

*useknutia*), za ktorou začína byť autokorelačná funkcia nulová (alebo zistiť, že také  $k_0$  neexistuje). Keďže pre konkrétnu radu teoretickú autokorelačnú funkciu  $\rho_k$  nepoznáme, je dôležité vedieť, ako blízko nule musí byť  $r_k$ , aby sme s určitou spoľahlivosťou mohli tvrdiť, že  $\rho_k = 0$ . Na to sa používa *Bartlettova aproximácia*, podľa ktorej ak je  $\rho_k = 0$  pre  $k > k_0$ , tak za predpokladu (asymptotickej) normality je

$$r_k \sim N\left(0, \frac{1}{n} \left(1 + 2 \sum_{j=1}^{k_0} r_j^2\right)\right) \text{ pre } k > k_0. \quad (4.1.8)$$

Popri autokorelačnej funkcii  $\rho_k$  sa používa tiež *parciálna autokorelačná funkcia*  $\rho_{kk}$  (PACF). Hodnota  $\rho_{kk}$  je definovaná ako parciálny korelačný koeficient medzi  $y_i$  a  $y_{i-k}$  pri pevných  $y_{i-k+1}, \dots, y_{i-1}$  (viď podkapitola 2.3). Na výpočet *odhadu parciálnej autokorelačnej funkcie* sa používa rekurentný vzťah uvedený v [2]

$$r_{11} = r_1, \quad r_{kk} = \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} \cdot r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} \cdot r_j} \text{ pre } k > 1, \quad (4.1.9)$$

kde

$$r_{kj} = r_{k-1,j} - r_{kk} \cdot r_{k-1,k-j} \text{ pre } j = 1, \dots, k-1. \quad (4.1.10)$$

Podobne ako ACF, tak aj PACF môže mať bod useknutia, a teda tiež slúži na identifikáciu modelu. V tomto prípade sa používa *Quenouilleova aproximácia*, podľa ktorej ak je  $\rho_{kk} = 0$  pre  $k > k_0$ , tak za predpokladu (asymptotickej) normality je

$$r_{kk} \sim N\left(0, \frac{1}{n}\right) \text{ pre } k > k_0. \quad (4.1.11)$$

## 4.2 Základné modely MA, AR a ARMA

Definujeme *lineárny proces* ako

$$\begin{aligned} y_i &= \varepsilon_i + \psi_1 \varepsilon_{i-1} + \psi_2 \varepsilon_{i-2} + \dots = (1 + \psi_1 B + \psi_2 B^2 + \dots) \varepsilon_i = \\ &= \psi(B) \varepsilon_i, \end{aligned} \quad (4.2.1)$$

kde  $\{\varepsilon_i\}$  je biely šum (definovaný v podkapitole 3.5) a  $B$  je operátor časového posunu, pričom platí

$$B^k \varepsilon_i = \varepsilon_{i-k}, \quad k = 1, 2, \dots. \quad (4.2.2)$$

Lineárny proces je *invertibilný* v prípade, že (4.2.1) môžeme vyjadriť v tvare

$$\begin{aligned} y_i &= \pi_1 y_{i-1} + \pi_2 y_{i-2} + \dots + \varepsilon_i, \text{ a teda} \\ \varepsilon_i &= y_i - \pi_1 y_{i-1} - \pi_2 y_{i-2} - \dots = \pi(B) y_i. \end{aligned} \quad (4.2.3)$$

Medzi základné modely Boxovej-Jenkinsovej metodológie patrí *proces kĺzavých súčtov rádu  $q$*  (MA( $q$ )), ktorý je v tvare

$$y_i = \varepsilon_i + \theta_1 \varepsilon_{i-1} + \dots + \theta_q \varepsilon_{i-q} = \theta(B) \varepsilon_i, \quad (4.2.4)$$

kde  $\{\varepsilon_i\}$  je biely šum,  $\theta_1, \dots, \theta_q$  sú parametre a  $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$  je *operátor kĺzavých súčtov*. Proces MA( $q$ ) je stacionárny s nulovou strednou hodnotou a rozptylom

$$\sigma_y^2 = (1 + \theta_1^2 + \dots + \theta_q^2) \sigma^2, \text{ kde } \sigma^2 = \text{var}(\varepsilon_i). \quad (4.2.5)$$

Autokorelačná funkcia tohto procesu má tvar

$$\rho_k = \begin{cases} \frac{\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \dots + \theta_q^2} & \text{pre } k = 1, \dots, q \\ 0 & \text{pre } k > q, \end{cases} \quad (4.2.6)$$

a teda bod useknutia  $k_0 = q$ . Naproti tomu PACF procesu MA( $q$ ) bod useknutia nemá, ale je obmedzená lineárnou kombináciou geometricky klesajúcich postupností a sínusoid s geometricky klesajúcimi amplitúdami. Proces MA( $q$ ) je invertibilný, ak všetky korene  $z_1, \dots, z_p$  polynómu  $\theta(z)$  ležia mimo jednotkového kruhu (tzn.  $|z_i| > 1, i = 1, \dots, q$ ).

Ďalším významným modelom je *autoregresný proces rádu  $p$*  (AR( $p$ )), ktorý má tvar

$$y_i = \varphi_1 y_{i-1} + \dots + \varphi_p y_{i-p} + \varepsilon_i, \text{ a teda} \quad (4.2.7)$$

$$y_i - \varphi_1 y_{i-1} - \dots - \varphi_p y_{i-p} = \varphi(B) y_i = \varepsilon_i,$$

kde  $\{\varepsilon_i\}$  je biely šum,  $\varphi_1, \dots, \varphi_p$  sú parametre a  $\varphi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$  je *autoregresný operátor*. Proces AR( $p$ ) je stacionárny, ak všetky korene  $z_1, \dots, z_p$  polynómu  $\varphi(z)$  ležia mimo jednotkového kruhu. V takom prípade má proces AR( $p$ ) nulovú strednú hodnotu a rozptyl

$$\sigma_y^2 = \frac{\sigma^2}{1 - \varphi_1 \rho_1 - \dots - \varphi_p \rho_p}. \quad (4.2.8)$$

Autokorelačná funkcia stacionárneho procesu AR( $p$ ) splňuje diferenčnú rovnicu

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \dots + \varphi_p \rho_{k-p} \text{ pre } k > 0. \quad (4.2.9)$$

Jej riešenie môžeme vyjadriť v tvare

$$\rho_k = \alpha_1 z_1^{-k} + \alpha_2 z_2^{-k} + \dots + \alpha_p z_p^{-k} \text{ pre } k \geq 0, \quad (4.2.10)$$

kde  $z_1, \dots, z_p$  sú navzájom rôzne korene polynómu  $\varphi(z)$  a  $\alpha_1, \dots, \alpha_p$  sú pevné koeficienty. ACF procesu AR( $p$ ) je lineárnou kombináciou klesajúcich geometrických postupností a sínusoid rôznych frekvencií s geometricky klesajúcimi amplitúdami.

Parciálna autokorelačná funkcia  $\rho_{kk}$  procesu  $AR(p)$  má bod useknutia  $k_0 = p$ . Proces  $AR(p)$  je vždy invertibilný.

Kombináciou procesu  $AR$  a  $MA$  dostaneme *zmiešaný proces rádiv*  $p$  a  $q$ , ktorý sa označuje ako  $ARMA(p,q)$  a jeho tvar je

$$y_i = \varphi_1 y_{i-1} + \dots + \varphi_p y_{i-p} + \varepsilon_i + \theta_1 \varepsilon_{i-1} + \dots + \theta_q \varepsilon_{i-q},$$

$$\text{tj. } \varphi(B)y_i = \theta(B)\varepsilon_i. \quad (4.2.11)$$

Podmienka stacionarity procesu  $ARMA(p,q)$  je rovnaká ako u procesu  $AR(p)$ , podmienka invertibility rovnaká ako u  $MA(q)$ . Stacionárny proces  $ARMA(p,q)$  má nulovú strednú hodnotu a jeho autokorelačná funkcia splňuje rovnicu

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \dots + \varphi_p \rho_{k-p} \quad \text{pre } k > q, \quad (4.2.12)$$

ktorej riešenie je v tvare

$$\rho_k = \alpha_1 z_1^{-k} + \alpha_2 z_2^{-k} + \dots + \alpha_p z_p^{-k} \quad \text{pre } k \geq \max(0, q - p + 1), \quad (4.2.13)$$

kde  $z_1, \dots, z_p$  sú navzájom rôzne korene polynómu  $\varphi(z)$ . ACF ani PACF procesu  $ARMA(p,q)$  nemajú bod useknutia.

### 4.3 Identifikácia modelu

Model budeme identifikovať na základe tvaru autokorelačnej a parciálnej autokorelačnej funkcie. Záver môžeme vyvodiť na základe nasledujúcej tabuľky.

	$AR(p)$	$MA(q)$	$ARMA(p,q)$
$\rho_k$	neexistuje $k_0$ ; $\rho_k$ v tvare krivky $U$	$k_0 = q$	neexistuje $k_0$ ; $\rho_k$ v tvare krivky $U$ po prvých $q - p$ hodnotách
$\rho_{kk}$	$k_0 = p$	neexistuje $k_0$ ; $\rho_{kk}$ obmedzená krivkou $U$	neexistuje $k_0$ ; $\rho_k$ obmedzená krivkou $U$ po prvých $p - q$ hodnotách

**Tab. 4.3.1.** Tvar ACF a PACF stacionárnych a invertibilných procesov  $AR$ ,  $MA$  a  $ARMA$  ( $U$  označuje krivku v tvare lineárnej kombinácie klesajúcich geometrických postupností a sínusoid s geometricky klesajúcimi amplitúdami)

Potencionálny bod useknutia  $k_0$  je možné testovať pomocou Bartlettovej aproximácie (4.1.8) s približným kritickým oborom (na hladine 5%)

$$|r_k| \geq 2 \sqrt{\frac{1}{n} \left( 1 + 2 \sum_{j=1}^{k_0} r_j^2 \right)} \quad \text{pre niektoré } k > k_0 \quad (4.3.1)$$

alebo pomocou Quenouilleovej aproximácie (4.1.11) s kritickým oborom (na hladine 5%)

$$|r_{kk}| \geq 2 \sqrt{\frac{1}{n}} \text{ pre niektoré } k > k_0. \quad (4.3.2)$$

Existencia štatisticky významných korelácií  $r_k$  (respektíve  $r_{kk}$ ) pre  $k > k_0$  vedie k zamietnutiu  $k_0$  ako bodu useknutia ACF (respektíve PACF). V prípade, že sa model nedá jednoznačne určiť pomocou tab. 4.3.1, je možné pre dané dáta odhadnúť viac modelov a vybrať z nich napríklad podľa optimalizačného kritéria AIC (*Akaike information criterion*) uvedenom v [8]

$$AIC(p, q) = \ln(\hat{\sigma}_{p,q}^2) + \frac{2(p+q)}{n}, \quad (4.3.3)$$

kde  $\hat{\sigma}_{p,q}^2$  je odhadnutý rozptyl bieleho šumu procesu ARMA( $p, q$ ). Preferujeme model s najnižšou hodnotou AIC.

#### 4.4 Odhad parametrov modelu

K odhadu parametrov v MA, AR a ARMA modeloch je možné použiť momentovú metódu, metódu najmenších štvorcov alebo metódu maximálnej vierohodnosti. Príslušné postupy sú popísané v knihách [2] a [5]. K ich praktickému prevedeniu sa používa špecializovaný software (NCSS, EViews, R, Wolfram Mathematica a iné).

#### 4.5 Overovanie modelu

Za overovanie modelu považujeme:

- 1) Kontrolu splnenia podmienky stacionarity. To znamená, že v AR a ARMA modeloch kontrolujeme, či korene odhadnutého autoregresného polynómu  $\varphi(z)$  ležia mimo jednotkového kruhu. Podobne v MA a ARMA modeloch kontrolujeme splnenie podmienky invertibility. Teda kontrolujeme, či korene odhadnutého polynómu kľzavých súčtov  $\theta(z)$  ležia mimo jednotkového kruhu.
- 2) Kontrolu vypočítaného bieleho šumu  $\{\hat{\varepsilon}_i\}$  pomocou grafického záznamu, korelogramu (nekorelovanosť), histogramu (normalita) a podobne.
- 3) Testovanie nekorelovanosti vypočítaného bieleho šumu  $\{\hat{\varepsilon}_i\}$  (za predpokladu normality) pomocou testu založeného na Bartlettovej aproximácii. Nulová hypotéza nekorelovanosti má kritický obor (na hladine 5%)

$$|r_k(\hat{\varepsilon}_i)| \geq 2 \sqrt{\frac{1}{n}}, \text{ pre } k = 1, 2, \dots \quad (4.5.1)$$



Ďalšou možnosťou je použitie *Q-testu*, ktorý testuje významnosť prvých  $K$  autokorelácií odhadnutého bieleho šumu (kde  $K$  sa doporučuje voliť vo výške  $K \approx \sqrt{n}$ ). Nulovou hypotézou je tvrdenie, že pôvodná rada sa riadi modelom  $\text{ARMA}(p, q)$ . Ako  $Q$ -štatistika sa často používa *Boxova-Pierceova štatistika* s kritickým oborom (na hladine  $\alpha$ )

$$Q = n \sum_{k=1}^K (r_k(\hat{\varepsilon}_i))^2 \geq \chi^2_{1-\alpha}(K - p - q), \quad (4.5.2)$$

kde  $\chi^2_{1-\alpha}(K - p - q)$  je  $(1 - \alpha)$ -kvantil  $\chi^2$  rozdelenia.

## 5 Aplikácia

### 5.1 Príklad 1

Máme k dispozícii hodnoty ročnej miery inflácie (v %) v USA od roku 1971 do roku 2008, čo predstavuje 38 pozorovaní. Označíme ich ako  $y = (y_{1971}, \dots, y_{2008})$ .

Ďalej máme pozorovania veličín (ročné hodnoty v období 1971 až 2008 v USA):

1. HDP (v dolároch)
2. Nárast HDP (v %)
3. Hodnota CPI
4. Import (v dolároch)
5. Import (v % z HDP)
6. Export (v dolároch)
7. Export (v % z HDP)
8. Reálna úroková miera (v %)
9. Peniaze (LCU)
10. Priemer indexu DJI (Dow Jones)
11. Nárast indexu DJI (v %)
12. Nárast populácie (v %)

Dáta boli spracované v programe *Wolfram Mathematica 8.0.1* (na kontrolu výsledkov boli použité programy *NCSS* a *R*).

Aby sme mali predstavu o veľkosti a variabilite dát, uvedieme najskôr hodnoty priemerov a výberových smerodajných odchýlok.

	priemer	výberová smerodajná odchýlka
Inf.	4,64948	2,92729
1.	6,23273E+12	3,98283E+12
2.	3,06113	2,01875
3.	64,0386	26,7265
4.	8,16797E+11	6,95287E+11
5.	11,3399	2,94351
6.	6,30671E+11	4,72383E+11
7.	9,33309	1,63394
8.	4,38799	2,56215
9.	8,69505E+11	4,03043E+11
10.	4524,97	4154,09
11.	0,0897918	0,129457
12.	1,03947	0,142529

Tab. 5.1.1. Priemery a smerodajné odchýlky dát

Ďalej sa pozrime na tabuľky s hodnotami výberových korelácií a výberových parciálnych korelácií.

	Inf.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
Inf.	1	-0,58	-0,33	-0,62	-0,51	-0,43	-0,53	-0,19	-0,14	-0,67	-0,55	-0,40	-0,19
1.	-0,58	1	-0,15	0,98	0,98	0,96	0,99	0,77	0,12	0,95	0,96	0,12	-0,01
2.	-0,33	-0,15	1	-0,15	-0,17	-0,17	-0,18	-0,32	0,03	-0,09	-0,09	0,22	0,02
3.	-0,62	0,98	-0,15	1	0,93	0,95	0,95	0,80	0,27	0,98	0,90	0,18	0,07
4.	-0,51	0,98	-0,17	0,93	1	0,96	0,99	0,75	0,05	0,88	0,96	0,09	-0,08
5.	-0,43	0,96	-0,17	0,95	0,96	1	0,96	0,83	0,21	0,90	0,91	0,09	-0,10
6.	-0,53	0,99	-0,18	0,95	0,99	0,96	1	0,82	0,11	0,92	0,95	0,14	0,03
7.	-0,19	0,77	-0,32	0,80	0,75	0,83	0,82	1	0,30	0,78	0,72	0,12	0,24
8.	-0,14	0,12	0,03	0,27	0,05	0,21	0,11	0,30	1	0,23	0,04	0,28	0,05
9.	-0,67	0,95	-0,09	0,98	0,88	0,90	0,92	0,78	0,23	1	0,87	0,21	0,19
10.	-0,55	0,96	-0,09	0,90	0,96	0,91	0,95	0,72	0,04	0,87	1	0,09	0,00
11.	-0,40	0,12	0,22	0,18	0,09	0,09	0,14	0,12	0,28	0,21	0,09	1	0,18
12.	-0,19	-0,01	0,02	0,07	-0,08	-0,10	0,03	0,24	0,05	0,19	0,00	0,18	1

**Tab. 5.1.2.** Výberové korelačné koeficienty

	Inf.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
Inf.	-1	0,32	-0,52	-0,41	0,08	0,34	-0,49	0,62	0,27	-0,18	-0,33	-0,11	0,29
1.	0,32	-1	0,39	0,91	0,70	-0,50	-0,02	-0,09	-0,60	0,25	0,86	-0,37	-0,31
2.	-0,52	0,39	-1	-0,54	-0,29	0,54	-0,01	0,01	0,35	0,09	-0,27	0,15	0,21
3.	-0,41	0,91	-0,54	-1	-0,61	0,56	0,02	0,10	0,72	0,10	-0,78	0,27	0,22
4.	0,08	0,70	-0,29	-0,61	-1	0,69	0,70	-0,55	0,16	-0,26	-0,57	0,28	0,07
5.	0,34	-0,50	0,54	0,56	0,69	-1	-0,39	0,38	-0,09	0,22	0,44	-0,18	-0,43
6.	-0,49	-0,02	-0,01	0,02	0,70	-0,39	-1	0,89	0,25	0,03	0,04	0,00	0,25
7.	0,62	-0,09	0,01	0,10	-0,55	0,38	0,89	-1	-0,24	0,02	0,09	0,05	-0,04
8.	0,27	-0,60	0,35	0,72	0,16	-0,09	0,25	-0,24	-1	-0,26	0,50	-0,08	-0,18
9.	-0,18	0,25	0,09	0,10	-0,26	0,22	0,03	0,02	-0,26	-1	-0,23	0,16	0,44
10.	-0,33	0,86	-0,27	-0,78	-0,57	0,44	0,04	0,09	0,50	-0,23	-1	0,23	0,31
11.	-0,11	-0,37	0,15	0,27	0,28	-0,18	0,00	0,05	-0,08	0,16	0,23	-1	-0,13
12.	0,29	-0,31	0,21	0,22	0,07	-0,43	0,25	-0,04	-0,18	0,44	0,31	-0,13	-1

**Tab. 5.1.3.** Koeficienty parciálnych korelácií

V tab. 5.1.2 vidíme, že inflácia má najvyššie hodnoty výberového korelačného koeficientu (v absolútnej hodnote) s veličinami peniaze (LCU) a hodnotou CPI. Naopak slabo korelovaná je s veličinami export (v % z HDP), reálna úroková miera (v %) a nárast populácie (v %).

V tab. 5.1.3 sú hodnoty výberových koeficientov parciálnych korelácií spočítaných pomocou škálovanej matice vzorcom (2.3.9). Inflácia má najvyššiu absolútnu hodnotu parciálnej korelácie s veličinou export (v % z HDP), kedy hodnota významne vzrástla oproti hodnote z tab. 5.1.2. Vzhľadom k tomu, že štruktúra vzájomných zá-

vislostí v množine veličín, ktoré máme k dispozícii je zložitá, má parciálna korelácia väčšiu vypovedaciu hodnotu o vplyve jednotlivých premenných na infláciu.

Našou úlohou bude modelovať mieru inflácie metódou lineárnej regresie. Uvažujme model so štyrmi regresormi.

#### (a) výber štvorice regresorov na základe hodnôt parciálnych korelácií

Na základe tab. 5.1.3 zvolíme 4 regresory, ktoré majú významnejšie absolútne hodnoty parciálnych korelácií s infláciou. Regresory zvolíme napríklad nasledovne:

$X_1$  ako nárast HDP (v %),

$X_2$  ako hodnotu CPI,

$X_3$  ako import (% z HDP),

$X_4$  ako export (% z HDP).

- **Korelačná analýza (a)**

Výberová korelačná matica vektora  $(Y, X_1, X_2, X_3, X_4)^T$  je v nasledujúcej tabuľke (jedná sa o submaticu tab. 5.1.2).

	y	x1	x2	x3	x4
y	1,0000	-0,3299	-0,6221	-0,4255	-0,1891
x1	-0,3299	1,0000	-0,1499	-0,1736	-0,3168
x2	-0,6221	-0,1499	1,0000	0,9490	0,8017
x3	-0,4255	-0,1736	0,9490	1,0000	0,8283
x4	-0,1891	-0,3168	0,8017	0,8283	1,0000

**Tab. 5.1.4.** Výberové korelačné koeficienty

V prvom stĺpci je vidieť, že miera inflácie je najviac korelovaná s hodnotou CPI, čo pravdepodobne spôsobuje fakt, že CPI sa používa na výpočet miery inflácie. Ale aj tak sú hodnoty korelácie prvého stĺpca pomerne nízke. Najviac korelovanou dvojicou je import (v % z HDP) s hodnotou CPI. Naproti tomu premenná  $X_1$  (nárast HDP) neukazuje výraznú koreláciu so žiadnou z ostatných premenných.

Výberový koeficient mnohonásobnej korelácie premennej  $Y$  s vektorom premenných  $X_1, X_2, X_3, X_4$  je 0,929294. Vysoká hodnota znamená, že závislosť miery inflácie od ostatných premenných je možné modelovať lineárnou funkciou  $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_4 X_4 + e$ .

Matica parciálnych korelácií:

	y	x1	x2	x3	x4
y	1,0000	-0,6024	-0,8834	0,7048	0,5554
x1	-0,6024	1,0000	-0,5022	0,4365	0,1225
x2	-0,8834	-0,5022	1,0000	0,9039	0,5332
x3	0,7048	0,4365	0,9039	1,0000	-0,1873
x4	0,5554	0,1225	0,5332	-0,1873	1,0000

**Tab. 5.1.5.** Koeficienty parciálnych korelácií

V tomto prípade sme pracovali len s množinou premenných  $Y, X_1, X_2, X_3$  a  $X_4$ , nájdená sa teda o submaticu tab. 5.1.3. Opäť v prvom stĺpci je vidieť hodnoty týkajúce sa miery inflácie. Je vidieť, že všetky štyri hodnoty sú pomerne vysoké, a teda jednotlivé závislosti nie sú zapríčinené pôsobením ostatných veličín.

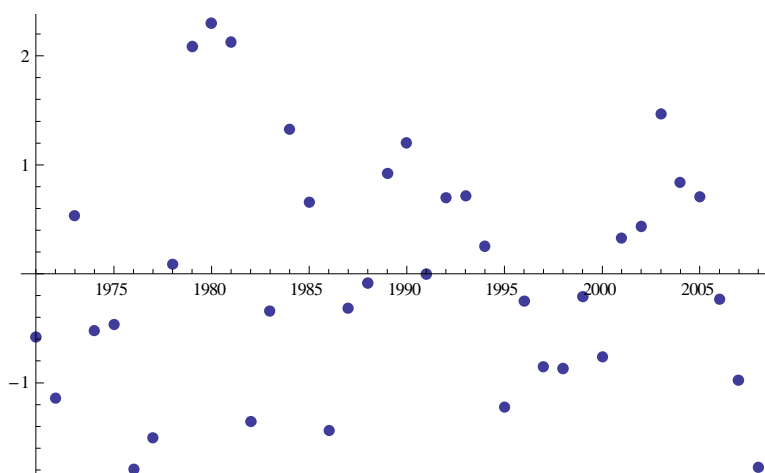
- **Regresná analýza (a)**

Vzťah medzi mierou inflácie a vektorom regresorov vyjadríme pomocou lineárnej funkcie. Vektor parametrov  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_4)^T$  odhadneme metódou najmenších štvorcov ako  $\mathbf{b} = (b_0, \dots, b_4)^T$  a dostaneme model

$$\hat{Y}_{(1)} = -0,2850 - 0,4336 X_1 - 0,2436 X_2 + 1,2399 X_3 + 0,8359 X_4. \quad (5.1.1)$$

Znamienka regresných koeficientov odpovedajú znamienkam parciálnych korelácií z tabuliek 5.1.3 a 5.1.5. Reziduálny súčet štvorcov  $S_{e(1)} = 43,2504$ . Koeficient determinácie  $R_{(1)}^2 = 0,863587$ , čo znamená, že model vysvetľuje viac ako 86% variability miery inflácie. Korigovaný koeficient determinácie  $\bar{R}_{(1)}^2 = 0,847052$ .

Hodnoty rezíduí (chýb) majú priemer  $-5,493 \cdot 10^{-15}$  a sú nepravidelne roztrúsené okolo nulovej úrovne, čo môžeme vidieť v nasledujúcom obrázku.



**Obr. 5.1.1.** Bodový graf rezíduí v závislosti na čase

Shapiro-Wilkovým testom normality nezamietame normalitu vektora chýb na hladine  $\alpha = 0,05$  ( $p\text{-hodnota} = 0,49018$ ).

Výsledky testu nulovosti regresných koeficientov sú zhrnuté v nasledujúcej tabuľke (test je založený na kritickom obore (3.4.3)).

regresor	odhad parametru	p-hodnota
x1	-0,4336	1,2886E-04
x2	-0,2436	2,1128E-12
x3	1,2399	2,2810E-06
x4	0,8359	5,3281E-04

**Tab. 5.1.6.** P-hodnoty testov nulovosti parametrov  $\beta_1, \dots, \beta_4$

Test nulovosti zamietá nulovosť  $\beta_i$  pre  $i = 1, \dots, 4$  na hladine  $\alpha = 0,05$ .

Na porovnanie ešte uvedieme, ako sa zmenia koeficienty determinácie v prípade, že z modelu vypustíme jeden z regresorov.

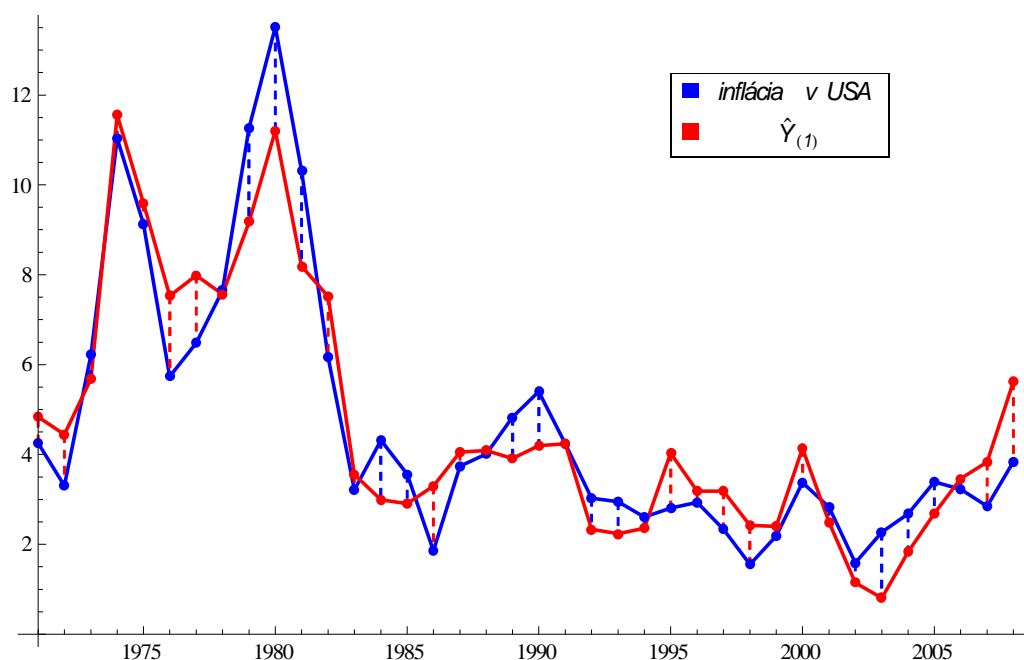
zjednodušenie modelu o regresor:	x1	x2	x3	x4
koeficient determinácie:	0,785908	0,378648	0,728907	0,802721
korigovaný koeficient determinácie:	0,767018	0,323823	0,704987	0,785314

**Tab. 5.1.7.** Koeficienty determinácie po vypustení regresorov

Vo všetkých štyroch prípadoch sa hodnoty výrazne znížili, čo znamená, že ďalšie zjednodušovanie modelu by už menej kvalitne popisovalo dáta. Všimnime si, že v prípade vynechania premennej  $X_2$  (hodnota CPI) je zníženie najväčšie, a to súhlasí aj s korelačnou analýzou a spôsobom výpočtu inflácie.

- **Záver (a)**

K modelovaniu inflácie môžeme použiť všetky štyri zvolené premenné, ktoré vysvetlia 86% jej variability pomocou modelu (5.1.1). V nasledujúcom obrázku môžeme vidieť vývoj skutočnej inflácie a nášho modelu  $\hat{Y}_{(1)}$ .



Obr. 5.1.2. Skutočné hodnoty inflácie a hodnoty modelu  $\hat{Y}_{(1)}$

Intuitívny výber štvorice regresorov však nemusí byť najvhodnejší, už len preto, že rôznych štvoríc pri 12-tich veličinách je 495.

#### (b) optimálny výber regresorov $X_1, X_2, X_3, X_4$

Na optimalizáciu výberu musíme zvoliť kritérium, podľa ktorého budeme vyberať jednotlivé regresory. Vhodné bude napríklad minimalizovať reziduálny súčet štvorcov  $S_e$  (a teda maximalizovať koeficient determinácie). Pri našich dátach teda nájdeme také vektory  $X_1, X_2, X_3, X_4$  (celkovo 495 možností), pri ktorých je po použití lineárnej regresie minimálny reziduálny súčet štvorcov (tento postup nazveme ako algoritmus (1b)). Túto úlohu vyriešime a dostaneme regresory

$X_1$  ako import (v dolároch),

$X_2$  ako export (v dolároch),

$X_3$  ako export (v % z HDP),

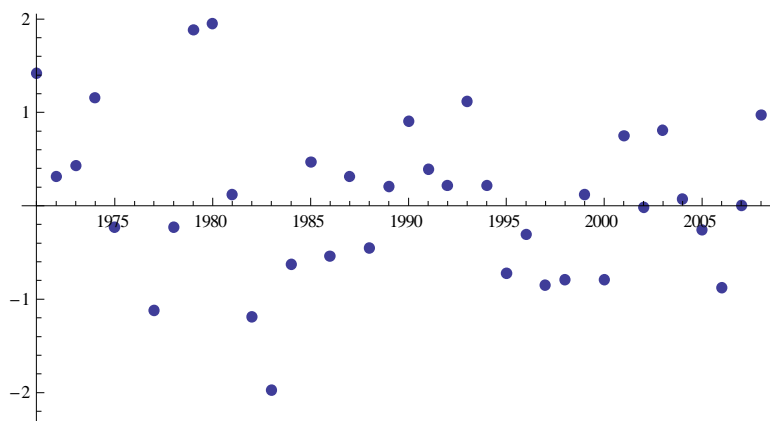
$X_4$  ako peniaze (LCU).

#### • Regresná analýza (b)

Lineárny model popisujúci mieru inflácie:

$$\begin{aligned} \hat{Y}_{(2)} = & -10,7556 + 1,5883 \cdot 10^{-11} X_1 - 2,9424 \cdot 10^{-11} X_2 \\ & + 2,8473 X_3 - 6,4231 \cdot 10^{-12} X_4, \end{aligned} \quad (5.1.2)$$

kde reziduálny súčet štvorcov  $S_{e_{(2)}} = 35,1717$  je hodnota, na základe ktorej sme regresory zvolili. Všimnime si, že hodnota klesla oproti hodnote  $S_{e_{(1)}}$  a teda  $\hat{Y}_{(2)}$  lepšie modeluje skutočnú mieru inflácie v USA. Potvrdzuje to aj koeficient determinácie  $R_{(2)}^2 = 0,889067$  a korigovaný koeficient determinácie  $\bar{R}_{(2)}^2 = 0,875621$ , ktoré vzrástli oproti hodnote v časti (a).



**Obr. 5.1.3.** Bodový graf rezíduí v závislosti na čase

Hodnoty rezíduí sú nepravidelne roztrúsené okolo nulovej úrovne s priemerom  $-7,25034 \cdot 10^{-14}$ . Shapiro-Wilkov test nezamieta normalitu vektora chýb na hladine  $\alpha = 0,05$  ( $p$ -hodnota = 0,472354).

regresor	odhad parametru	p-hodnota
x1	1,5883E-11	2,5569E-07
x2	-2,9425E-11	4,0112E-07
x3	2,8473	4,6165E-12
x4	-6,4231E-12	6,5744E-06

**Tab. 5.1.8.** P-hodnoty testov nulovosti parametrov  $\beta_1, \dots, \beta_4$

Test nulovosti parametrov  $\beta_i$  pre  $i = 1, \dots, 4$  zamietajú možnosť zjednodušovať model na hladine  $\alpha = 0,05$  (viď tab. 5.1.8).

Zmeny koeficientov determinácie v prípade vypustenia jedného z regresorov z modelu (5.1.2) vidíme v nasledujúcej tabuľke.

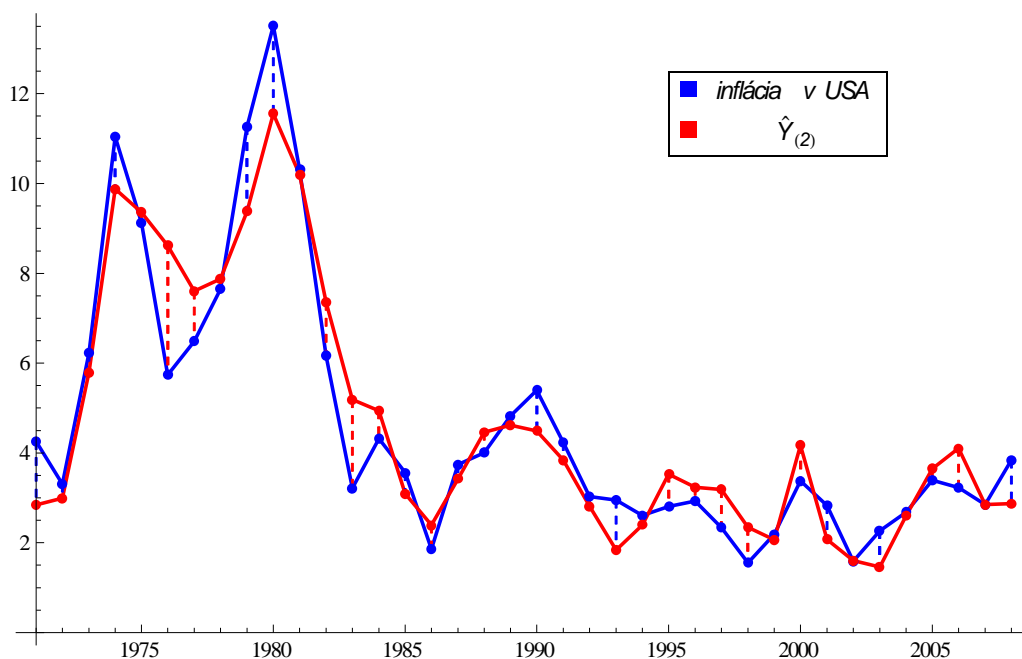
zjednodušenie modelu o regresor:	x1	x2	x3	x4
koeficient determinácie:	0,749073	0,755683	0,517903	0,792864
korigovaný koeficient determinácie:	0,726932	0,734126	0,475365	0,774587

**Tab. 5.1.9.** Koeficienty determinácie po vypustení regresorov



- **Záver (b)**

Model (5.1.2) vysvetľuje takmer 89% variability miery inflácie, čo je maximálna možná hodnota pri použití štyroch regresorov z daných dát. Zaujímavé však je, že takýto model vyžaduje ako premenné hodnoty exportu (v dolároch) aj exportu (v % z HDP). V nasledujúcom obrázku vidíme ako sa modelu  $\hat{Y}_{(2)}$  darí modelovať skutočnú mieru inflácie.



**Obr. 5.1.4.** Skutočné hodnoty inflácie a hodnoty modelu  $\hat{Y}_{(2)}$

### Záver príkladu 1

V oboch prípadoch sa mieru inflácie v USA podarilo modelovať pomocou vhodne zvolených regresorov. V prípade, že by sme neobmedzili počet premenných, viedlo by nás to k použitiu čo najväčšieho počtu regresorov, čo je spôsobené tým, že ani korigovaný koeficient determinácie nedokáže dostatočne penalizovať veľký počet regresorov.

Na záver uvedieme tabuľku, ktorá ukazuje úspešnosť modelu pri použití algoritmu (1b) s 5, 6 a 7 regresormi.

Počet regresorov:	5	6	7
Regresory (očíslované podľa zadania)	2, 3, 4, 6, 7	2, 3, 4, 5, 6, 7	2, 3, 4, 5, 6, 7, 11
Reziduálny súčet štvorcov:	28,0602	26,5954	25,1237
Koeficient determinácie:	0,9115	0,9161	0,9208
Korigovaný koeficient determinácie:	0,8977	0,8999	0,9023

**Tab. 5.1.10.** Optimálne použitie viacerých regresorov

V prípade so šiestimi regresormi, pomocou testu nulovosti na hladine  $\alpha = 0,05$ , nezamietame nulovosť parametra pri regresore 5 (import (v % z HDP)). Pri siedmich regresoroch nezamietame nulovosť parametrov pre regresory 5 a 11 (nárast indexu DJI v %). A teda v oboch prípadoch model zjednodušujeme na model s piatimi rovnakými regresormi:

$X_1$  ako nárast HDP (v %),

$X_2$  ako hodnotu CPI,

$X_3$  ako import (v dolároch),

$X_4$  ako export (v dolároch),

$X_5$  ako export (v % z HDP).

Model s uvedenými regresormi je nasledovný:

$$\hat{Y}_{(3)} = -8,054 - 0,2548 X_1 - 0,113 X_2 + 1,718 \cdot 10^{-11} X_3 - 3,019 \cdot 10^{-11} X_4 + 2,756 X_5. \quad (5.1.3)$$

## Zhrnutie

Zistili sme, že s použitím 4 alebo 5 regresorov môžeme popísať približne 90% variability miery inflácie. V modeli s vyšším počtom regresorov a teda vyšším koeficientom determinácie sa už objavujú štatisticky nevýznamné regresné koeficienty.

Sústredíme sa ďalej na výsledný optimálny model s 5 regresormi (5.1.3). Hodnoty rezíduí sú nepravidelne roztrúsené okolo nulovej úrovne s priemerom  $-6,92312 \cdot 10^{-14}$ . Normalitu vektora chýb nezamietame na hladine  $\alpha = 0,05$  pomocou Shapiro-Wilkovho testu normality ( $p\text{-hodnota} = 0,76802$ ).

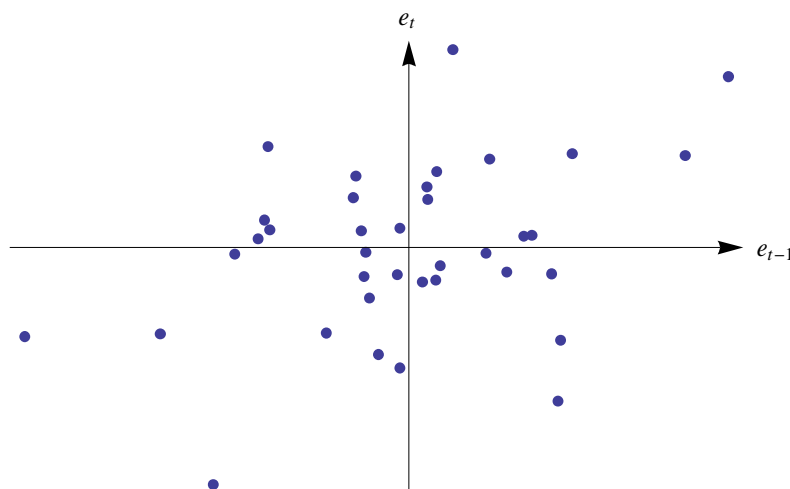
regresor	odhad parametru	p-hodnota
x1	-0,2548	4,9210E-03
x2	-0,1129	2,5556E-06
x3	1,7180E-11	3,2280E-09
x4	-3,0190E-11	1,6230E-08
x5	2,756	5,0662E-12

**Tab. 5.1.11.** P-hodnoty testov nulovosti parametrov  $\beta_1, \dots, \beta_5$

V tab. 5.1.11 vidíme, že testy nulovosti (na hladine  $\alpha = 0,05$ ) pre jednotlivé parametre vylučujú ďalšie zjednodušovanie modelu.

Na zistenie autokorelovanosti rezíduí použijeme Durbinov-Watsonov test podľa vzorca (3.5.3) a dostaneme testovú štatistiku  $DW = 1,2331$  (odhad parametra  $\varphi$  zo vzorca (3.5.1) je  $\hat{\varphi} = 0,3703$ ). Zo štatistických tabuliek získame kritické hodnoty

$d_L = 1,2042$  a  $d_U = 1,7916$ , z čoho vyplýva, že test je bez vypovedacej hodnoty (viď obr. 3.5.1). Podobne ani z obr. 5.1.5 nedokážeme vyvodiť záver.



**Obr. 5.1.5.** Bodový graf rezíduí v závislosti na oneskorených hodnotách

Model (5.1.3) zrejme vykazuje multikolinearitu, pretože výberové korelačné koeficienty medzi dvojicami regresorov vo viacerých prípadoch presahujú hodnotu 0,9 (viď tab. 5.1.12).

	x1	x2	x3	x4	x5
x1	1,0000	-0,1499	-0,1668	-0,1764	-0,3168
x2	-0,1499	1,0000	0,9295	0,9532	0,8017
x3	-0,1668	0,9295	1,0000	0,9873	0,7486
x4	-0,1764	0,9532	0,9873	1,0000	0,8234
x5	-0,3168	0,8017	0,7486	0,8234	1,0000

**Tab. 5.1.12.** Výberové korelačné koeficienty (hodnoty nad 0,9 sú vyznačené červenou farbou)

Rovnako aj  $C_p = \|X^T X\| \cdot \|(X^T X)^{-1}\| = 3,3231 \cdot 10^{26} \gg 1$ , a teda matica  $X^T X$  je zle podmienená a nájdenie jej inverzie je numericky náročné, čo naznačuje multikolinearitu (problém zle podmienenej matice a s tým spojené numerické chyby hlási aj samotný software). V súvislosti s multikolinearitou sa pozrime aj na tabuľku hodnôt faktorov variančnej inflácie (ich vysoké hodnoty potvrdzujú efekt multikolinearity).

	x1	x2	x3	x4	x5
VIF	1,2235	11,8073	92,4098	153,334	7,58191

**Tab. 5.1.13.** Faktory variančnej inflácie

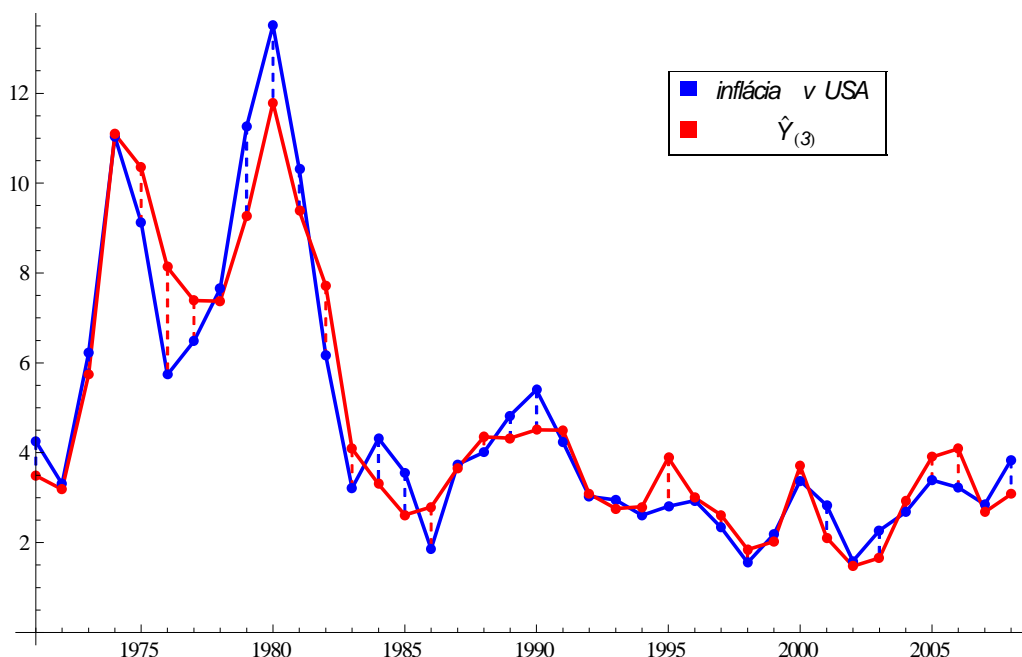
V nasledujúcej tabuľke vidíme, ako sa zmení presnosť modelu v prípade, že vypustíme jeden z regresorov.

zjednodušenie modelu o regresor:	x1	x2	x3	x4	x5
koeficient determinácie:	0,8863	0,8215	0,7313	0,7567	0,5994
korigovaný koeficient determinácie:	0,8725	0,7999	0,6987	0,7272	0,5509

**Tab. 5.1.14.** Koeficienty determinácie po vypustení regresorov

V prípade, že vypustíme regresor  $X_1$  nie je pokles koeficientu determinácie vysoký, avšak tento regresor nebol podozrivý zo spôsobovania multikolinearity (viď hodnoty korelačných koeficientov regresora  $X_1$  s ostatnými regresormi v tab. 5.1.12 a hodnotu faktora variančnej inflácie pre  $X_1$  v tab. 5.1.13). V ostatných prípadoch sa koeficient determinácie zníži výraznejšie. Ak by sme však chceli eliminovať multikolinearitu museli by sme vypustiť viac ako jeden regresor a tým by sme výrazne znehodnotili model. Z toho dôvodu budeme multikolinearitu ignorovať.

Na záver sa pozrime na obrázok modelu (5.1.3).



**Obr. 5.1.6.** Skutočné hodnoty inflácie a hodnoty modelu  $\hat{Y}_{(3)}$

V CD prílohe je priložený program s názvom *program.nb*, ktorý po zvolení regresorov užívateľom zobrazí požadovaný model (celý postup vrátane algoritmu (1b) je v programe *projekt.nb*). Vyžaduje spustenie v programe Wolfram Mathematica 8.

## 5.2 Príklad 2

V tomto príklade použijeme hodnoty ročnej miery inflácie (v %) v USA už od roku 1961 po rok 2009. Označíme ich ako  $y = (y_{1961}, \dots, y_{2009})$ . Vývoj inflácie v čase budeme modelovať metódami analýzy časových rád, ktoré sme popísali v kapitole 4.

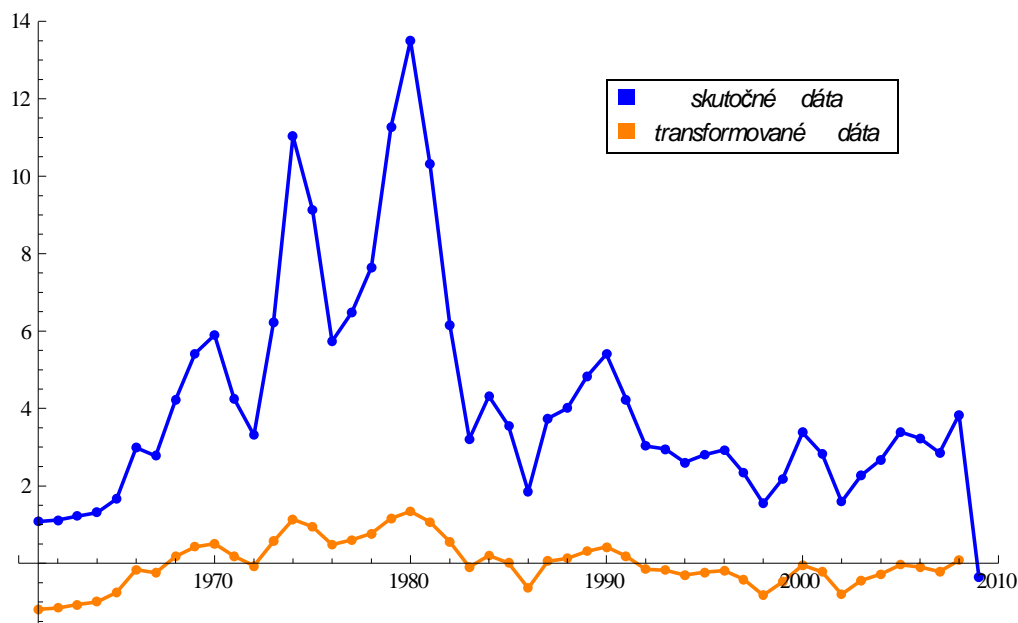
rok	inflácia	rok	inflácia	rok	inflácia	rok	inflácia	rok	inflácia
1961	1,0752	1971	4,2559	1981	10,316	1991	4,235	2001	2,8262
1962	1,1161	1972	3,3056	1982	6,1606	1992	3,0288	2002	1,586
1963	1,2141	1973	6,2201	1983	3,2124	1993	2,9517	2003	2,2701
1964	1,3086	1974	11,035	1984	4,3173	1994	2,6074	2004	2,6772
1965	1,6685	1975	9,1319	1985	3,5611	1995	2,8054	2005	3,3927
1966	2,991	1976	5,737	1986	1,8587	1996	2,9312	2006	3,2259
1967	2,7756	1977	6,4865	1987	3,7409	1997	2,3377	2007	2,8527
1968	4,2177	1978	7,6475	1988	4,0091	1998	1,5523	2008	3,8391
1969	5,4147	1979	11,266	1989	4,827	1999	2,188	2009	-0,356
1970	5,8953	1980	13,509	1990	5,398	2000	3,3769		

Tab. 5.2.1. Ročné miery inflácie (v %) v USA

- **Identifikácia a odhad modelu**

K zaisteniu stacionarity a nulovej strednej hodnoty transformujeme dáta logaritmovaním a odčítaním priemeru. Avšak hodnota inflácie v roku 2009 je záporná (deflácia), preto skrátime radu o posledný člen (túto hodnotu neskôr použijeme na overenie predpovede). Transformované dáta označíme ako  $z = (z_{1961}, \dots, z_{2008})$ .

V nasledujúcom obrázku môžeme vidieť grafy skutočných ( $y$ ) a transformovaných ( $z$ ) dát.

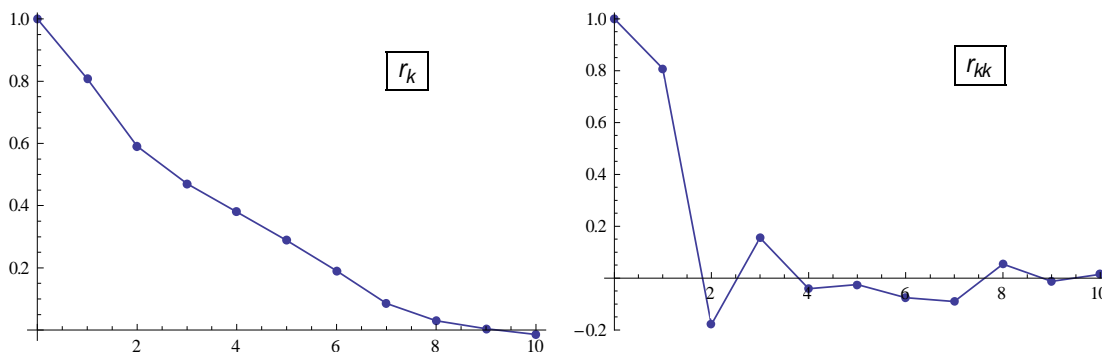


Obr. 5.2.1. Skutočné hodnoty inflácie a hodnoty transformovaných dát

Ďalej sa pozrime na odhad autokorelačnej a parciálnej autokorelačnej funkcie pre radu  $z$ .

	1	2	3	4	5	6	7	8	9	10
ACF	0,808	0,591	0,47	0,38	0,289	0,191	0,086	0,03	0,005	-0,014
PACF	0,808	-0,179	0,156	-0,041	-0,025	-0,076	-0,09	0,054	-0,013	0,014

Tab. 5.2.2. Odhadnutá ACF a PACF rady  $z$



Obr. 5.2.2. Odhadnutý korelogram a parciálny korelogram rady  $z$

S využitím tab. 4.3.1 sa ako najvhodnejší model javí proces AR(1) ( $r_k$  je klesajúca a  $r_{kk}$  má pravdepodobne bod useknutia  $k_0=1$ ). Pre kontrolu s využitím Quenouillovej aproximácie je podľa (4.3.2)

$$|r_{kk}| < 2 \sqrt{\frac{1}{n}} = 2 \sqrt{\frac{1}{48}} = 0,2887 \text{ pre } k > 1. \quad (5.2.1)$$

Parameter modelu AR(1) môžeme odhadnúť metódou najmenších štvorcov a dostaneme  $\varphi_1 = 0,808205$ .

Ďalšou možnosťou identifikácie je použitie optimalizačného kritéria AIC (viď podkapitola 4.3) v prípade, že sme pre dané dáta odhadli viac konkrétnych modelov. Na to nám posluží vstavaná funkcia programu Wolfram Mathematica 8 s názvom *Hannan Rissanen Estimate*. Pomocou tejto funkcie dostaneme 5 modelov ARMA( $p,q$ ), kde  $p, q = 0,1,2,3$  (hodnoty 5 a 3 sú nami zvolené vstupné parametre).

	AR(1)	ARMA(1,1)	ARMA(1,2)	ARMA(1,3)	MA(3)
AIC	-2,27163	-2,25328	-2,25925	-2,20348	-1,71699

Tab. 5.2.3. Modely navrhnuté procedúrou Hannan Rissanen Estimate a ich hodnoty AIC

Minimálna hodnota AIC je v prípade zvolenia modelu AR(1), a teda touto metódou volíme ten istý model

$$z_i = 0,808205 z_{i-1} + \varepsilon_i. \quad (5.2.2)$$

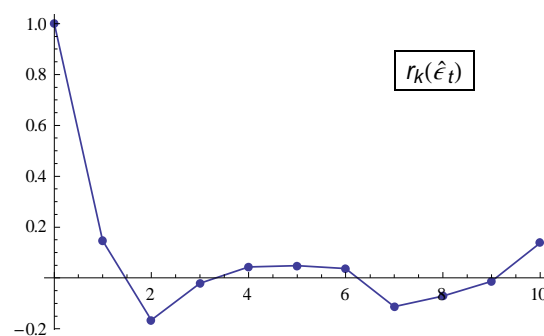
- **Overovanie modelu**

K overeniu podmienky stacionarity zistíme, či korene odhadnutého autoregresného polynómu ležia mimo jednotkový kruh. V našom prípade koreň polynómu  $\varphi(B) = (1 - 0,808205 B)$  je  $B_1 = 1,23731 > 1$  a teda odhadnutý model AR(1) splňuje stacionaritu.

Hodnoty rezíduí majú priemer 0,00137. V tab. 5.2.4 a na obr. 5.2.3 vidíme odhadnutú ACF rezíduí.

	1	2	3	4	5	6	7	8	9	10
ACF	0,147	-0,165	-0,021	0,044	0,048	0,037	-0,112	-0,07	-0,013	0,138

**Tab. 5.2.4.** Odhadnutá ACF rezíduí



**Obr. 5.2.3.** Odhadnutý korelogram rezíduí

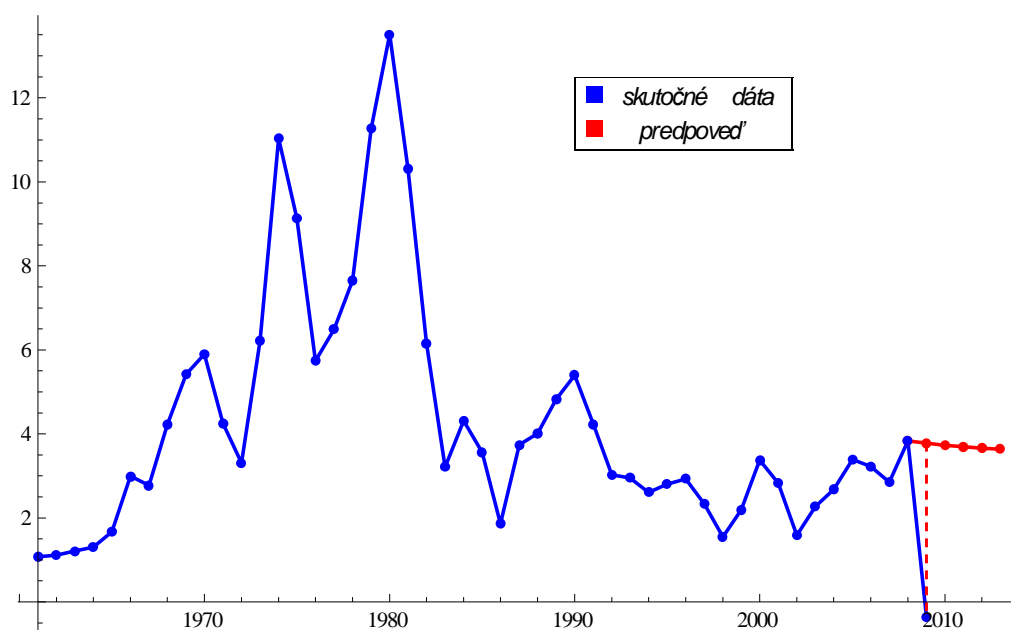
Nekorelovanosť rezíduí overíme pomocou (4.5.1). Pretože

$$|r_k(\hat{\varepsilon}_i)| < 2 \sqrt{\frac{1}{n}} = 2 \sqrt{\frac{1}{48}} = 0,2887, \text{ pre } k = 1, 2, \dots, 10, \quad (5.2.3)$$

nezamietame nulovú hypotézu nekorelovanosti na hladine významnosti 5%.

- **Predpoveď**

Na predpoveď rady  $z$  s využitím zvoleného modelu AR(1) použijeme opäť vstavanú funkciu programu Wolfram Mathematica 8 s názvom *Best Linear Predictor*, pomocou ktorej dostaneme hodnoty piatich predpovedí. Následne tieto predpovede transformujeme na predpovede rady  $y$  (hodnoty ročnej miery inflácie). Podľa predpovedí nastane mierny pokles v hodnotách, čo môžeme vidieť v nasledujúcom obrázku.



**Obr. 5.2.4.** Skutočné hodnoty inflácie a hodnoty predpovedí (prerušovaná čiara je znázorňovaný rozdiel medzi predpoveďou a skutočnou hodnotou v roku 2009)

Na začiatku príkladu sme hodnotu  $y_{2009} = -0,356$  vynechali (z dôvodu logaritmovania). Podľa predpovede však hodnota ročnej miery inflácie v roku 2009 má hodnotu 3,7795, čo znamená, že ani predpoveď s využitím vhodného ARMA modelu nedokáže predpovedať taký výrazný pokles. Tento pokles mohol byť spôsobený oživením ekonomiky po prekonaní finančnej krízy z rokov 2008 až 2009.



## Záver

V tejto práci sme spracovali teóriu korelačnej analýzy, regresnej analýzy a analýzy jednorozmerných časových rád. Následne sme teóriu aplikovali v dvoch príkladoch na modelovanie miery inflácie v USA.

V prvom prípade sme mali k dispozícii 12 veličín (možných regresorov) použiteľných na model lineárnej regresie. Zo začiatku sme sa obmedzili na model so štyrmi regresormi. V prvej časti sme štvoricu regresorov vyberali na základe hodnôt parciálnych korelácií, pričom takto zvolený model vysvetlil 86% variability miery inflácie. V druhej časti sme pomocou optimálneho výberu regresorov modelom vysvetlili až 89% variability miery inflácie. V závere príkladu sme sa zaoberali presnejším modelom s použitím piatich regresorov, na ktorom sme testovali autokorelovanosť rezíduí a multikolinearitu. Efekt autokorelovanosti rezíduí dokázaný nebol, multikolinearita sa v skúmaných dátach prejavila.

V druhom prípade sme pomocou Boxovej-Jenkinsovej metodológie hľadali vhodný ARMA model na popis vývoja inflácie v čase a predpoveď jej budúceho chovania. Na základe autokorelačnej a parciálnej autokorelačnej funkcie sme zvolili model AR(1). S využitím funkcie Best Linear Predictor v programe Wolfram Mathematica sme predpovedali vývoj miery inflácie na 5 rokov. Ukázalo sa, že model AR(1) nie je schopný predpovedať výrazné zmeny v priebehu časovej rady, pretože hodnota prvej predpovede sa oproti reálnej hodnote výrazne líšila.

# Literatúra

- [1] ANDEL, Jiří. *Základy matematické statistiky*. Praha : MATFYZPRESS, 2007. 358 s.
- [2] CIPRA, Tomáš. *Finanční Ekonometrie*. Praha : EKOPRES, 2008. 538 s.
- [3] DUPAČOVÁ, Jitka; HURT, Jan; ŠTĚPÁN, Josef. *Stochastic Modeling in Economics and Finance*. Dordrecht : Kluwer, 2002. 392 s.
- [4] KLEINBAUM, David G.; KUPPER, Lawrence L. *Applied Regression Analysis and Other Multivariable Methods*. North Scituate, Massachusetts : DUXBURY PRESS, 1978. Another way of describing partial correlations, s. 166.
- [5] WEI, William W.Ss. *Time Series Analysis. Univariate and Multivariate Methods*. New York : ADDISON-WESLEY, 1990. 478 s.
- [6] WHITTAKER, Joe. *Graphical models in applied multivariate statistics*. New York : Wiley, 1990. 466 s.
- [7] ZICHOVÁ, Jitka. *Plánování experimentů a predikční vícerozměrová analýza*. Praha : Karolinum, 2007. 134 s.
- [8] *Reference and User's Guide to Time Series Pack for Mathematica*. Champaign : Wolfram Research, 1995. Estimation of Correlation Function and Model Identification, s. 81-82.
- [9] ČVUT, FJFI [online]. [cit. 2011-07-18]. Numerická matematika. Dostupné z WWW: <<http://kfe.fjfi.cvut.cz/~klimo/nm/l4.pdf>>.
- [10] *Stanford University* [online]. [cit. 2011-07-18]. Durbin-Watson Critical Values. Dostupné z WWW: <<http://www.stanford.edu/~clint/bench/dwcrit.htm>>.
- [11] *Český statistický úřad / ČSÚ* [online]. [cit. 2011-07-18]. Míra inflace. Dostupné z WWW: <[http://www.czso.cz/csu/redakce.nsf/i/mira\\_inflace](http://www.czso.cz/csu/redakce.nsf/i/mira_inflace)>.
- [12] *Český statistický úřad / ČSÚ* [online]. [cit. 2011-07-18]. KDYŽ SE ŘEKNE INFLACE, resp. MÍRA INFLACE. Dostupné z WWW: <[http://www.czso.cz/csu/redakce.nsf/i/kdyz\\_se\\_rekne\\_inflace\\_resp\\_mira\\_inflace#1](http://www.czso.cz/csu/redakce.nsf/i/kdyz_se_rekne_inflace_resp_mira_inflace#1)>.
- [13] Inflation. In *Wikipedia : the free encyclopedia* [online]. Wikipedia Foundation, [cit. 2011-07-18]. Dostupné z WWW: <<http://en.wikipedia.org/wiki/Inflation>>.
- [14] *Data | The World Bank* [online]. [cit. 2011-07-18]. Financial Sector. Dostupné z WWW: <<http://data.worldbank.org/topic/financial-sector>>.

## Prílohy

Na priloženom CD sa nachádzajú nasledujúce dokumenty

- Bakalárka práca (v pdf)
- Dáta použité v príkladoch
- Zdrojové kódy výpočtov pre jednotlivé príklady (v programe Wolfram Mathematica 8)